THE STUDY OF RISK FACTORS DEPENDENCE OF CONTEXT USING COMPATIBILITY FUNCTIONS

Cristina Gena Dascălu*, Georgeta Zegan**, Smărăndița Georgeta Păcuraru***

* Medical Informatics and Biostatistics Department, Faculty of Dental Medicine
**Orthodontics Department, Faculty of Dental Medicine
***Periodontology Department, Faculty of Dental Medicine
University of Medicine and Pharmacy „Gr.T. Popa", Iaşi, Romania

cdascalu@umfiasi.ro

Abstract: The risk factors analysis is a very important domain of interest in medicine, with significant implications in the diseases prevention programs. The methods to analyze a single risk factor's influence in a disease are well known; a special case is met when more than one risk factors interact in the same disease's manifestation. For this situation, often met in the medical practice, we need special tools to study the way in which the risks change; in our paper we propose such a tool – a compatibility function between risk factors, used to characterize the type of interactions between them which can appear. This function is built for the case of simultaneously action of two risk factors and than it is generalized for the case of N risk factors action. We also present some useful properties of this function and its significations related to the risk factors interaction. In the end, these theoretical notions are exemplified in a practical study about the nourishment habits influence over the dental mobility of 2389 people from Iasi – the risks are calculated using the classic method, as well as the new proposed compatibility function.

Introduction

By epidemiological studies that take place at a certain moment in time, denoted by t_0, we can emphasize different risk factors that influence a disease's development. A risk factor's association with a disease can be characterized by defining a corresponding magnitude. When the disease's context is modified, at a new time moment $t_0 + \Delta t$, it is possible to notice a different probability for the disease's development in the presence of that risk factor – because the risk factor passes into a new class of influence. Thus, the relations between risk factors are different because of their context, which can bring, for example, a new risk factor for the analyzed disease, which wasn't included before in the study.

We propose in this paper a solution to study the risk factors context dependence – a method to update the probability to get a disease in the presence of a risk factor according to the context created by the presence of other factors.

Materials and methods

Let F be a set of N factors, denoted by F_1, F_2, F_3, ... F_N, whose association with a certain disease is biologically plausible and it was demonstrated for each risk factor F_i separately [1]. Let $\Omega = \{\omega_i\}$ be the set of magnitude classes which characterize the association degree between the risk factors and the studied disease.

Example: Ω can contain 5 classes, defined by the following predicates:

- "F_i is a major positive factor";
- "F_i is a moderate positive factor";
- "F_i is a weak positive factor";
- "F_i is an indifferent factor";
- "F_i is an negative factor".

We know the probabilities $P_k(i)$ for a risk factor F_i belonging to the magnitude class ω_k. By changing the relation between factors according with the context, the probabilities $P_k(i)$ are also modified. Let's take for two risk factors F_i and F_j, the hypothesis: F_i belongs to the magnitude class ω_k, and F_j belongs to the magnitude class ω_l.

In order to characterize the compatibility degree between these two hypothesis we define a positive function $C_{kl}(i, j)$, having the calculation formula:

$$C_{kl}(i, j) = \frac{\Pr(F_i \in \omega_k \mid F_j \in \omega_l)}{\Pr(F_i \in \omega_k)} = \frac{\Pr(F_i \in \omega_k \cap F_j \in \omega_l)}{\Pr(F_i \in \omega_k) \cdot \Pr(F_j \in \omega_l)}$$

(the probabilities are estimated using the frequencies distributions).

The function $C_{kl}(i, j)$ defined in this manner is a measure of the compatibility between the hypothesis "F_i is a risk factor having the magnitude ω_k" and "F_j is a risk factor having the magnitude ω_l"; the function has a bigger value when the two hypothesis are compatible, and a smaller value when the two hypothesis are not compatible. We call this function "compatibility function between the factors F_i and F_j".

ISSN: 1727-1983

This function has a few specific properties and significations, which allow interpreting the nature of the risk factor interaction.

Properties:

1. The compatibility function is simmetrically, namely $C_{kl}(i, j) = C_{lk}(j, i)$.

2. The $C_{kl}(i, j)$ function values range, $Val(C_{kl}(i, j)) = \{0,\infty\}$.

Notice: $C_{kl}(i, j)$ tends to overweight the compatibility situation. In order to avoid this difficulty, we propose the following transformations of the compatibility function:

a) $C_{kl}^1(i, j) = \log (C_{kl}(i, j))$, $C_{kl}^1(i, j)$ = the logarithmic expression of $C_{kl}(i, j)$;

b) Let $\overline{C}_{kl}(i, j)$ be the normalized expression of $C_{kl}(i, j)$, defined in such a way so $Val(\overline{C}_{kl}(i, j)) = \{0,1\}$. We define $C_{kl}^2(i, j) = \log(\overline{C}_{kl}(i, j))$.

3. The $C_{kl}^1(i, j)$ function values range, $Val(C_{kl}^1(i, j)) = \{-\infty,\infty\}$.

4. The $C_{kl}^2(i, j)$ function values range, $Val(C_{kl}^2(i, j)) = \{-1,1]$.

The *significations* of the compatibility function defined above are:

- The values $C_{kl}^1(i, j) > 0$, respectively $C_{kl}^2(i, j) \in \{-1, 0\}$ define situations of synergy or compatibility (the simultaneous action of the two factors leads to the amplification of their cumulative effect, which is higher that each factor's effect, taken separately):

$$C_{kl}^1(i, j) > 0 \Leftrightarrow \log (C_{kl}(i, j)) > 0$$
$$\Leftrightarrow C_{kl}(i, j) > 1$$
$$\Leftrightarrow \frac{\Pr(F_i \in \omega_k \cap F_j \in \omega_l)}{\Pr(F_i \in \omega_k) \cdot \Pr(F_j \in \omega_l)} > 1$$
$$\Leftrightarrow \Pr(F_i \in \omega_k \cap F_j \in \omega_l) > \Pr(F_i \in \omega_k) \cdot \Pr(F_j \in \omega_l)$$

- The values $C_{kl}^1(i, j) = 0$, respectively $C_{kl}^2(i, j) = 0$ define situations of independence: the cumulative effect of the two factors is equal with the product of their separate effects:

$$C_{kl}^1(i, j) = 0 \Leftrightarrow C_{kl}(i, j) = 1$$
$$\Leftrightarrow \Pr(F_i \in \omega_k \cap F_j \in \omega_l) = \Pr(F_i \in \omega_k) \cdot \Pr(F_j \in \omega_l)$$

- The values $C_{kl}^1(i, j) < 0$, respectively $C_{kl}^2(i, j) \in \{0, 1]$ define situations of antagonism, or incompatibility (the cumulative effect of the two factors is smaller than each factor's effect, taken separately):

$$C_{kl}^1(i, j) < 0 \Leftrightarrow C_{kl}(i, j) < 1$$
$$\Leftrightarrow \Pr(F_i \in \omega_k \cap F_j \in \omega_l) < \Pr(F_i \in \omega_k) \cdot \Pr(F_j \in \omega_l)$$

Being given the F_i factor, for each factor F_j related with F_i, where $j \neq i$, we define a credibility coefficient $cc_k(i, j)$ of the F_i ownership to ω_k as it follows:

$$cc_k(i, j) = \sum_{l=1}^{N} C_{kl}(i, j) \cdot P_l(j)$$

Considering the way we defined the compatibility function $C_{kl}(i, j)$, it follows that the credibility coefficient $cc_k(i, j)$ will have a high value when the compatibility and $P_l(j)$ are high and will decrease when either the compatibility is small, either $P_l(j)$ is small, either both expressions are small. In other words, the credibility of the assumption "F_j factor belongs to the magnitude class ω_k" in the context created by the presence of the F_i factors grows as much as the context is more compatible and the compatible factors have a higher probability of association with the disease.

Let $P_k^t(i)$ be the probabilities of a F_i factor apartenence to the magnitude class ω_k, at the t moment of time. The credibility coefficient induced to F_i by F_j at the t moment of time, $cc_k^t(i, j)$ is calculated with the formula:

$$cc_k^t(i, j) = \sum_{l=1}^{N} C_{kl}(i, j) \cdot P_l^t(j)$$

This credibility coefficient can be used to update the probabilities at the time moment $t + \Delta t$, namely to obtain the probabilities $P_k^{t+\Delta t}(i)$ (this aspect will make the subject of some further researches).

We can also make a *generalization* of the compatibility function, available in the case of simultaneous presence of more than two risk factors for a disease. Let's take N risk factors, F_1, F_2, F_3, ... F_N, whose association with a certain disease is biologically plausible and it was demonstrated for each F_i factor separately. Let's also take $\Omega = \{\omega_l\}$, as the set of magnitude classes which characterize the association degree between the risk factors and the studied disease, and $P_k(i) = \Pr(F_i \in \omega_k)$ the probability of the assumption "F_i risk factor belongs to the ω_k magnitude class".

Let's take for all the N factors F_{i1}, F_{i2}, … F_{iN} the hypothesis: $F_{ij} \in \omega_{kj}$, $j \in \{1, 2, \ldots N\}$. We study the compatibility between these N hypothesis – namely the way in which a risk factor's presence in a disease can be influenced by the simultaneous presence of other risk factors in the same disease. We will use also a positive function, $C_N(i_1, i_2, \ldots i_N)$, defined as it follows:

$$C_N(i_1, i_2, \ldots i_N) = \frac{\Pr\left(\bigcap_{j=1}^{N} F_{ij} \in \omega_{kj}\right)}{\prod_{j=1}^{N} \Pr(F_{ij} \in \omega_{kj})}$$

The function $C_N(i_1, i_2, \ldots i_N)$ is called compatibility function of the $F_{i1}, F_{i2}, F_{i3}, \ldots F_{iN}$ factors.

Properties:

1. The compatibility function $C_N(i_1, i_2, \ldots i_N)$ doesn't depend by the $F_{i1}, F_{i2}, F_{i3}, \ldots F_{iN}$ factors order of action.

2. The $C_N(i_1, i_2, \ldots i_N)$ function values range, $Val(C_N(i_1, i_2, \ldots i_N)) = \{0,\infty\}$.

In order to eliminate, as in the previous case, the phenomenon of compatibility overweighting, we can use the compatibility function transformation:

$$C_N^1(i_1, i_2, \ldots i_N) = \log (C_N(i_1, i_2, \ldots i_N)),$$

which also has the property:

3. The $C_N^1(i_1, i_2, \ldots i_N)$ function values range, $Val(C_N^1(i_1, i_2, \ldots i_N)) = \{-\infty,\infty\}$.

The *significations* of the compatibility function $C_N^1(i_1, i_2, \ldots i_N)$ are the following (similarly with the case of two factors compatibility):

- The values $C_N^1(i_1, i_2, \ldots i_N) > 0$ define synergy, or compatibility situations;

ISSN: 1727-1983

- The values C_N^1 $(i_1, i_2, \ldots i_N) = 0$ define independence situations;
- The values C_N^1 $(i_1, i_2, \ldots i_N) < 0$ define antagonism situations.

Results

We calculated the compatibility function in order to characterize the risk factors interaction, in a practical study made on a set of 2389 people from Iasi county about the nourishment habits and the way in which these habits act as risk factors for a certain dental disease – the dental mobility. The studied nourishment habits were: the weekly consumption of fresh vegetables, the weekly consumption of fresh fruits, the weekly consumption of milk, cheese, meal, and the daily consumption of coffee, sweets, and juices (the data were achieved from the database of the WHO Center for Romania – Dental Health Section).

We calculate first the individual risks for each nourishment habit separately, using the data contingency tables [2, 3]. The data contingency tables are given in the Table 1, and the individual risks are given in Table 2.

Table 1: The data contingency tables for the analyzed sample

	Dental mobility		TOTAL		Dental mobility		TOTAL
	Present	Absent			Present	Absentă	
Fresh vegetables consumption / week				Meal consumption / week			
NO (never)	29	150	179	NO (never)	6	50	56
YES (4 or more times / week)	32	334	366	YES (4 or more times / week)	97	786	883
Fresh fruits consumption / week				Daily coffee consumption			
NO (never)	21	83	104	YES	99	856	955
YES (4 or more times / week)	55	577	632	NO	182	1080	1262
Milk consumption / week				Daily juices consumption			
NO (never)	44	261	305	YES	99	1067	1166
YES (4 or more times / week)	92	671	763	NO	177	867	1044
Cheese consumption / week				Daily sweets consumption			
NO (never)	22	179	201	YES (3 or more times / week)	30	336	366
YES (4 or more times / week)	47	333	380	NO (never)	103	410	513

Table 2: The calculated individual risks for the analyzed sample – using the classic method

RISK FACTOR	EXPOSED SUBJECTS		UNEXPOSED SUBJECTS		S_1 – Chance of disease at exposed subjects	S_2 – Chance of disease at unexposed subjects	RR – relative risk of disease
	R_1 – risk of disease presence	R_2 – risk of disease absence	R_3 – risk of disease presence	R_4 – risk of disease absence	Odds ratio = S_1 / S_2		
Low consumption of fresh vegetables / week	**0.162**	0.838	**0.087**	0.913	0.193	0.096	1.853
					o.r. = 2.01		
Low consumption of fresh fruits / week	**0.202**	0.798	**0.087**	0.913	0.253	0.095	2.320
					o.r. = 2.66		
Low consumption of milk / week	**0.144**	0.856	**0.121**	0.879	0.169	0.137	1.196
					o.r. = 1.23		
Low consumption of cheese / week	**0.109**	0.891	**0.124**	0.876	0.123	0.141	0.885
					o.r. = 0.87		
Low consumption of meal / week	**0.107**	0.893	**0.110**	0.890	0.120	0.123	0.975
					o.r. = 0.98		
Coffee consumption	**0.104**	0.896	**0.144**	0.856	0.116	0.168	0.719
					o.r. = 0.69		
Juices consumption	**0.085**	0.915	**0.170**	0.830	0.092	0.204	0.501
					o.r. = 0.45		
High consumption of sweets / day	**0.082**	0.918	**0.200**	0.800	0.089	0.251	0.408
					o.r. = 0.35		

In order to identify the risk factors, we have to compare the risks of disease apparition at exposed subjects vs. unexposed subjects. Also, in order to establish the magnitude class for each risk factor, we use the relative risk of disease apparition, according to the following classification:

- $RR \leq 1$ – indifferent factor;
- $RR \in (1.00, 1.50]$ – weak positive factor;
- $RR \in (1.50, 2.00]$ – moderate positive factor;
- $RR \geq 2.00$ – major positive factor.

(this classification was chosen according with the particular nature of the sample and the relative risks size range – with the same results we can use instead of relative risks, the odds ratios).

Notice: The relative risks were calculated using the formula:

ISSN: 1727-1983

RR = R_1 (the risk of disease presence at exposed subjects) / R_3 (the risk of disease presence at unexposed subjects)

According to these considerations, we can conclude that the following nourishment habits can be defined as risk factors for the dental mobility apparition:

- the low weekly consumption of fresh vegetables – moderate positive factor;
- the low weekly consumption of fresh fruits – major positive factor;
- the low weekly consumption of milk – weak positive factor.

The other habits don't act as risk factors in dental mobility, being characterized as indifferent factors – that is why we will not include them in the further analyses about compatibility.

The next step is to analyze the interaction between the identified risk factors. The risk factors are denoted as it follows:

- low consumption of fresh vegetables – F_1;
- low consumption of fresh fruits – F_2;
- low consumption of milk – F_3.

We start with the multiple data tables which show the observed cases for each combination of risk factors and we calculate the risks of disease, the relative risks and the interaction indexes T on an additive and a multiplicative scale [2, 4] (the obtained results are presented in Table 3a, b). In this way we obtain a first characterization of the interaction nature between risk factors, and a primary identification of the synergy or antagonism situations.

Notice: In Table 3a, b, in order to analyze the risk factors interaction, we used the following notation and calculation formulas:

Let's take the case of interaction between two risk factors, F_1 and F_2. We denote by F_{11}, respectively F_{21} the risk factors presence, and by F_{10} and F_{20} their absence. The disease risks will be defined as:

$R_{11} = pr(B+|F_{11} \& F_{21})$: the probability (risk) to contact the disease in the presence of both factors;

$R_{01} = pr(B+|F_{10} \& F_{21})$: the probability (risk) to contact the disease in the absence of F_1 and the presence of F_2;

$R_{10} = pr(B+|F_{11} \& F_{20})$: the probability (risk) to contact the disease in the presence of F_1 and the absence of F_2;

$R_{00} = pr(B+|F_{10} \& F_{20})$: the probability (risk) to contact the disease in the absence of both factors.

The relative risks are calculated using the relations:

$RR_{11} = R_{11}/R_{00}$ – the relative risk to contact the disease in the presence of both factors;

$RR_{10} = R_{10}/R_{00}$ – the relative risk to contact the disease in the presence of F_1;

$RR_{01} = R_{01}/R_{00}$ – the relative risk to contact the disease in the presence of F_2.

Based on the relative risks described above, we can calculate the interaction index T between risk factors on an additive scale, using the formula:

$$T = (R_{11} - R_{10}) + (R_{00} - R_{01}).$$

In a similar way we can calculate the interaction index T on a multiplicative scale, using the formula:

$$T = (R_{11} * R_{00}) / (R_{10} * R_{01}).$$

(these interaction indexes provide the classical method to measure and to analyze the processes that appear in case of multiple risk factors action).

Table 3a: The analysis of interaction between two risk factors

	Dental mobility		TOTAL	Disease risk	Relative risks	The interaction index T
	Present	Absent				
F_1 & F_2						
F_{11} & F_{21}	11	33	44	$R_{11}^{12} = 0.250$	$RR_{11}^{12} = 4.750$	Additive scale:
F_{11} & F_{20}	1	18	19	$R_{10}^{12} = 0.053$	$RR_{10}^{12} = 1.000$	T = 0.028
F_{10} & F_{21}	2	7	9	$R_{01}^{12} = 0.222$	$RR_{01}^{12} = 4.222$	Multiplicative scale:
F_{10} & F_{20}	15	270	285	$R_{00}^{12} = 0.053$		T = 1.125
						F_1, F_2 : positive interaction
F_1 & F_3						
F_{11} & F_{31}	13	35	48	$R_{11}^{13} = 0.271$	$RR_{11}^{13} = 4.313$	Additive scale:
F_{11} & F_{30}	7	30	37	$R_{10}^{13} = 0.189$	$RR_{10}^{13} = 3.012$	T = 0.113
F_{10} & F_{31}	1	31	32	$R_{01}^{13} = 0.031$	$RR_{01}^{13} = 0.498$	Multiplicative scale:
F_{10} & F_{30}	13	194	207	$R_{00}^{13} = 0.063$		T = 2.877
						F_1, F_3 : positive interaction
F_2 & F_3						
F_{21} & F_{31}	3	28	31	$R_{11}^{23} = 0.097$	$RR_{11}^{23} = 1.251$	Additive scale:
F_{21} & F_{30}	4	17	21	$R_{10}^{23} = 0.190$	$RR_{10}^{23} = 2.462$	T = – 0.081
F_{20} & F_{31}	5	72	77	$R_{01}^{23} = 0.065$	$RR_{01}^{23} = 0.839$	Multiplicative scale:
F_{20} & F_{30}	26	310	336	$R_{00}^{23} = 0.077$		T = 0.605
						F_2, F_3 : negative interaction

ISSN: 1727-1983

Table 3b: The analysis of interaction between all the three risk factors

	Dental mobility		TOTAL	Disease risk	Relative risks	The interaction index T
	Present	Absent				
F_1 & F_2 & F_3						
F_{11} & F_{21} & F_{31}	3	11	14	$R_{111}^{123} = 0.214$	$RR_{111}^{123} = 3.921$	Additive scale*: T = – 0.335
F_{11} & F_{21} & F_{30}	2	5	7	$R_{110}^{123} = 0.286$	$RR_{110}^{123} = 5.229$	
F_{11} & F_{20} & F_{31}	0	4	4	$R_{101}^{123} = 0.000$	$RR_{101}^{123} = 0.000$	
F_{11} & F_{20} & F_{30}	1	5	6	$R_{100}^{123} = 0.167$	$RR_{100}^{123} = 3.050$	
F_{10} & F_{21} & F_{31}	0	2	2	$R_{011}^{123} = 0.000$	$RR_{011}^{123} = 0.000$	
F_{10} & F_{21} & F_{30}	1	2	3	$R_{010}^{123} = 0.333$	$RR_{010}^{123} = 6.100$	
F_{10} & F_{20} & F_{31}	1	26	27	$R_{001}^{123} = 0.037$	$RR_{001}^{123} = 0.678$	
F_{10} & F_{20} & F_{30}	10	173	183	$R_{000}^{123} = 0.055$		
					F_1, F_2, F_3 : negative interaction	

Based on the calculated relative risks and using the same classes of magnitude for the risk association with the disease, we obtain the following conclusions about the combined action of the studied risk factors (Table 4):

Table 4: The magnitude classes for combinations of risk factors

F_1 & F_2	$RR_{11}^{12} = 4.750$	major positive factor
F_1 & F_3	$RR_{11}^{13} = 4.313$	major positive factor
F_2 & F_3	$RR_{11}^{23} = 1.251$	weak positive factor
F_1 & F_2 & F_3	$RR_{111}^{123} = 3.921$	major positive factor

After this, we determine the compatibility degree between the identified risk factors, based on their model of context dependence.

We defined also 4 classes of magnitude for the risk factors association with the disease, denoted further by:

ω_1 – indifferent factor;
ω_2 – weak positive factor;
ω_3 – moderate positive factor;
ω_4 – major positive factor.

The probability for a risk factor 's belonging to a certain magnitude class is determined using its classical definition, the number of cases where the disease and the risk factor were both found and the sample's global size. In this way we obtain the following values:

$P(F_1 \in \omega_3) = 0.053$
$P(F_2 \in \omega_4) = 0.029$
$P(F_3 \in \omega_2) = 0.041$
$P(F_1\&F_2 \in \omega_4) = 0.031$
$P(F_1\&F_3 \in \omega_4) = 0.040$
$P(F_2\&F_3 \in \omega_2) = 0.006$
$P(F_1\&F_2\&F_3 \in \omega_4) = 0.012$

The hypothesis compatibility functions, according to the definition we proposed before, have the values presented in Table 5, and leads to the corresponding conclusions about the risk factors compatibility:

Table 5: The compatibility functions values

$C_{34}(1, 2) = 20.295$	$C_{34}^1(1, 2) = 1.307$	**synergy**
$C_{32}(1, 3) = 18.303$	$C_{32}^1(1, 3) = 1.263$	**synergy**
$C_{42}(2, 3) = 5.488$	$C_{42}^1(2, 3) = 0.739$	**synergy**
$C_{342}(1, 2, 3) = 194.967$	$C_{342}^1(1, 2, 3) = 2.290$	**synergy**

In this way we detect clearly the situations of compatibility or not compatibility between factors.

Conclusions

The compatibility function proposed in this paper can be used with good results in practical studies about the risk factors; it can be calculated easily, and its interpretation shows the nature of risk factors interaction in a clear and simple way.

Acknowledgements

We bring our special thanks to the WHO Collaborating Center from Iasi – Oral Health Section, which works next to the Faculty of Dental Medicine and especially to its Director, Prof. Vasile Burlui, M.D., Ph.D., for allowing us to use the center's medical databases as a practical support for our study about risk factors.

References

[1] LOFTUS, G.R., LOFTUS, E.L. (1988): 'Essence of Statistics', (Alfred A. Knopf, Inc.)
[2] OSTLE, B., MALONE, L.C. (1988): 'Statistics in Research', (Iowa State University Press / AMES)
[3] REMINGTON, R.D., SCHORK, M.A. (1985): 'Statistics with Applications to the Biological and Health Sciences', (Prentice – Hall, Inc.)
[4] RIZZI, A., VICHI, M., BOCK, H.H. (1998): 'Advances in Data Science and Classification', (Springer Verlag)

ISSN: 1727-1983