Anais da Academia Brasileira de Ciências (2011) 83(3): 1131-1146
(Annals of the Brazilian Academy of Sciences)
Printed version ISSN 0001-3765 / Online version ISSN 1678-2690
www.scielo.br/aabc

Correspondence model of occupational accidents

JUAN C. CONTE[1], EMILIO A. RUBIO[1], ANA I. GARCÍA[1] and FRANCISCO J. CANO[2]

[1]Cátedra de Bioestadística, Facultad de Medicina, Universidad de Zaragoza, C/ Domingo Miral, s/n, 50009-Zaragoza, España
[2]Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas,
Universidad Complutense de Madrid, Plaza de las Ciencias, no. 3. 28040-Madrid, España

Manuscript received on July 16, 2009; accepted for publication on December 1, 2010

ABSTRACT

We present a new generalized model for the diagnosis and prediction of accidents among the Spanish workforce. Based on observational data of the accident rate in all Spanish companies over eleven years (7,519,732 accidents), we classified them in a new risk-injury contingency table (19×19). Through correspondence analysis, we obtained a structure composed of three axes whose combination identifies three separate risk and injury groups, which we used as a general Spanish pattern. The most likely or frequent relationships between the risk and injuries identified in the pattern facilitated the decision-making process in companies at an early stage of risk assessment. Each risk-injury group has its own characteristics, which are understandable within the phenomenological framework of the accident. The main advantages of this model are its potential application to any other country and the feasibility of contrasting different country results. One limiting factor, however, is the need to set a common classification framework for risks and injuries to enhance comparison, a framework that does not exist today. The model aims to manage work-related accidents automatically at any level.

Key words: correspondence model, contingency analysis, risk, injury, occupational accidents.

INTRODUCTION

The aim of this paper is to present a generalized model of occupational accidents at a national scale, specifically for the Spanish workforce, by considering accidents as a compound risk-injury event. We aim to establish relationships of affinity between the component methods of these two variables, thereby generating a pattern against which Spanish companies can be analyzed. Although the model presented is multi-sectored, the same methodology, patterns can be applied to identify specific patterns for each individual industrial sector (metals, transport, chemicals, etc), if our goal is a more specific analysis of companies.

Risk assessment is currently an essential tool in managing safety in the workplace (Amendola 2002), and is used for predicting accidents among the workforce (Kjellén and Sklet 1995). This evaluation process involves three phases, all of them fundamental in the subsequent preventive action: identification, assessment and prioritization (Van Duijne et al. 2008, Frijters and Swuste 2008). In occupational accidents, the three mentioned phases take into consideration the error associated with the subjectivity required by the methodology used in the evaluation, although we need to redirect this methodology towards more objective models (Leveson 2004). This article aims to approach the problem in relation to the three phases above mentioned.

The standard evaluation of safety among the workforce begins by identifying an "assumed" (hypothetical) risk in the workplace, either by means of free observation or by means of a formal checklist (Rouhiainen 1992). Firstly, the evaluation uses tables of values to try to identify the probability and consequence variables

---

Correspondence to: Juan Carlos Conte Solano
E-mail: jcconte@unizar.es

at a general level (Fine 1973). Secondly, the risks are prioritized according their importance.

The main handicap in the standard evaluation of the "assumed" risks identified in the various jobs in a specific company lies in the fact that these risks are isolated, independent events, which may or may not affect individuals (Conte et al. 2007). To characterize an accident based on a risk before it actually happens is of little use at present, as it is subject to the fundamental premise of uncertainty. Therefore, once the "assumed" risk has been identified, it is not possible to prove either when the injury will occur or if it will occur, or even its level of severity.

The mentioned change of scale, moving from the rate of accidents (population) to the actual accident (the individual), adds a high level of randomness to the evaluation techniques used: the classic criterion variables play no proven role in the identification of risk (Körvers and Sonnemans 2008). Consequently, only individual, technical criterion prevails in the choice of the specific risk value.

In this sense, the conceptual generalization of risk proposed by Giddens (1994) and Beck (1999) is relevant, which they consider to be the modern focus of the forecasting and control of future (undesired) consequences of human actions, also combining two elements that have always been mutually exclusive so far: nature and society. They use the latter association to characterize the present day society, which they call a "risk society". According to these authors, a type of society is developing that will manage to overcome the problem of uncertainty generated by human actions or "manufactured uncertainty" (Giddens 1994) instead of by risks that can be forecast based on certain laws of science and natural systems (Beck 1999).

They see pre-industrial hazards conceived as "events of destiny", but that nowadays "industrial risks" pose the problem of a demand for social responsibility (accountability), enabling the assessment of risks that have not taken place to become the subject of prevention, compensation and expectation of preventive measures. Therefore, in Spain, 40% of companies have accidents. It is calculated that, in 60% of cases, the preventive risk action lacks any apparent use as it attempts to control a problem that does not exist.

Trying to minimize the uncertainty caused by the assessment techniques used, this article suggests the use of a more deterministic alternative: we propose to base the risk assessment of the control of occupational accidents upon the analysis of specific accidents suffered by a company, and not upon a set of "assumed" risks whose analysis goes beyond reality.

Therefore, this evaluation begins with an overall assessment of the risks that have actually taken place in a given workforce, avoiding the concept of their individuality, which is physically associated with spatial (Nicholson 1998), temporal (Sari et al. 2009) and material uncertainty (Hammer 1994).

Our aim, therefore, is to identify the real risks from a log of accidents and summarize them in a contingency table; we shall reach the criteria needed for their assessment and prioritization by the mathematical-statistical analysis.

This approach defines new quantifiable accident properties that, at least, help us to view the problem objectively, providing new control criteria based on a deeper knowledge about them.

Therefore, we have initially defined a general accident model (Conte et al. 2008), typical of a country, whose properties are applicable to any company within this country, the model is to be understood and used as a pattern, yardstick or standard for contrast (Garcia et al. 2009). We call it "acsom", the acronym of "accident soma" (accident body). To obtain it, we used a log, which is a temporal series of the accidents that occurred in this country, and each component in the series was reported in a risk-injury contingency table, which summarizes the accident rate recorded in each annual period.

Conceptually, acsom represents an "equilibrium diagram" of accidents. As we considered each accident as a compound event derived from the risk-injury pair, we identified the risk-injury (RI) type, which each accident comprises. The collection of these RI pairs for any given country constitutes its acsom-G, and it is presented as a compensated outline of its accident. It covers all the productive sectors, that is, all the positive and negative typological anomalies that characterize each area of activity. When we put these anomalies together, they combine with one another, compensating one another.

This produces a matrix (RI) diagram and marginal profiles (R or I), which are used as a standard of equilibrium/balance. The local patterns, or acsom-S, belonging to each branch of production, are interpreted in the same way.

The isolated profiles, which are marginal to the contingency table and which can be obtained for each company, in comparison to the marginal profiles in acsom-G, show deviations of accidents of the company, thereby allowing us to identify which types follow the R or I equilibrium profile, and which types deviate above or below it.

By means of a correspondence analysis, we present a global model (acsom-G) for accidents, its underlying data structure was made up of three groups of risks and injuries. We identified these three groups by colors (red, yellow and green) to recognize them visually. The colors do not indicate the level of seriousness of each group, but the features associated with the frequency of occurrence.

The correspondence model has been applied by various authors to the study of occupational accidents, although they have used it for the study of specific cases. Laflamme et al. (1991) applies the typology of the accidents in a Canadian car company and in a transport company. Williamson et al. (1996) analyzed 1738 industrial accidents in Australia to discover the relationship between work activity carried out in the workplace where the accidents happened and their nature. Baril et al. (2003) applied this methodology to a population of 13,728 injured people in order to establish the relationship among the activities of the injured workers, the types of injury suffered and the way the company deals with casualties.

The correspondence analysis was selected (Benzécri 1992, Greenacre and Blasius 1994) as the most suitable method to optimize the initial matrix functions, it reduced the information contained in the contingency table and established affinity relationships among the variable components of the table, thus obtaining a classification based on factorial coordinates (Joaristi and Lizasoaín 1999). Moreover, to obtain models of accidents in the companies, the correspondence analysis is undoubtedly the most suitable method because of its great power and elasticity. It makes no difference whether the table is finished or unfinished, that is to say, whether it presents structural or sample zeros, since $[\Sigma\lambda_i = \chi^2/N]$: the association grade among the variables defined by their eigenvalues (λ_i = eigenvalues from the diagonalization of the matrix; χ^2 = Pearson chi square value of the contingency table; N = total frequency of the table).

The contingency table obtained (Table I) shows three key elements: the total value, the marginal profiles, and the central body of the table or matrix. Each of these identified elements can be analyzed separately by using different methodologies.

In summary, the method is to compare the features of the accidents of a specific company as opposed to the features in their pattern of reference (acsom-G or S). This methodology is applicable to any company, regardless of size. In addition, one can automatically obtain the following: forecasts and predictions of accidents, and prioritizations of risks and injuries. Finally, it enables a follow-up in real time by implementing adequate control resources.

MATERIALS

In our model, an accident in the workforce is considered a compound event, composed of risks (R) and injuries (I). The risk is understood as a basic generating and component unit of the accident, which refers to the *physical process* inducing the injury. This latter, as it appears as the material evidence of one or more risks, is the basic compositional element or *biological product*, from which the occurrence of an accident on the individual is identified.

We have taken into consideration all the reports on occupational accidents notified over eleven years (7,519,732 accidents), registered (Ministerial Order 16-12-1987, BOE 311, of 29th December) and published by the Spanish Ministry of Labor (Secretaría General Técnica, Subdirección General de Estadísticas Sociales y Laborales). The risks and injuries mentioned in these notifications and reports are codified following the criteria of the International Labor Organization presented in the X International Conference of Statistical Labor of 1962. The data obtained are summarized in a contingency table of 19 risks (R) by 19 injuries (I), titled the starting risk-injury matrix.

The 19 categories of each variable (R or I) have a disjunctive and exhaustive codification. We considered the categories of the two initial multiple nominal variables as binary nominal variables, thereby obtaining 38 variables.

As the chosen variables R and I have a generic and exhaustive character, each one can be further subdivided, if it is useful, into other more specific derivatives of the main variable of reference.

To obtain an affinity relationship model among the variable components of acsom-G, we carried out the analysis on the average year matrix (Table I). It must be understood that, in order to obtain annual contingency tables, a Poisson type design (Aguilera 2001) was followed: the original frequencies that form the boxes of the contingency table for each period are independent random variables with Poisson distribution, and are filled freely. This basic table, or average year, of fictitious values appears as an incomplete table, which does not verify the hypothesis of symmetry of population probabilities and shows heterogeneity in its marginal distributions. It defines a theoretical body of yearly accidents, which allows the analytical determination of an acsom-G pattern.
The list of codes is as follows:

Risks	**Injuries**
Fall of Persons	**L1** Fractures
R1 Falls from a height (at different level)	**L2** Luxations
R2 Slips, trips, falls (at one or same level)	**L3** Twists, sprains, strains
Falls of Objects (on persons)	**L4** Back pain (lumbago, etc.)
R3 by Collapse	**L5** Slipped disc
R4 by Handling	**L6** Concussion and internal traumatism
R5 by Detachment	**L7** Amputations and loss of the eyeball
Treading on Objects	**L8** Other injuries (piercing, cutting, etc.)
R6 Pierces, cuts, etc.	**L9** Superficial traumatism
Colliding of persons against Objects (crashing)	**L10** Bruises, contusions and crushing
R7 Colliding against stationary objects	**L11** Objects in the eyes (particles, etc.)
R8 Colliding against mobile objects	**L12** Conjunctivitis
R9 Being struck by objects or tools (hits, cuts, etc.)	**L13** Poisoning and intoxications
R10 Projection of fragments or particles	**L14** Burns
R11 Catching by or between objects	**L15** Environmental hazard
R12 Accidents by mobile machinery and traffic	**L16** Suffocation
R13 Overstrain	**L17** Electric effects (shock, etc.)
R14 Exposure to thermal contact, heat strain	**L18** Radiation poisoning
R15 Exposure to electrical contact	**L19** Multiple injuries
R16 Exposure to chemical substances	
R17 Exposure to radiation	
R18 Explosions and fires	
R19 Accidents caused by living beings	

METHODS

The risk-injury matrix (Table I) was initially analyzed by using various skills of unsupervised learning techniques of data mining analysis (Hand et al. 2001), that is to say, by using statistical exploratory multivariate techniques, with the aim of identifying variable groups and verifying the obtained results, so as to define a global pattern (global model or acsom-G). Similarly, local patterns have been obtained (local model or acsom-S) for each branch of activity, outside the scope of this paper, to reflect the features that are typical of the industrial area to which it refers.

When we used the rate of accidents among the workforce the profiles in the contingency table follow Binomial and Poisson probability compound models (Rubio 1983) interpretable in its set as multinomial distributions (Aguilera 2006), Figures 1-2. Although an ideal representative option would be a bar diagram, a polygon is used in order to better see the differences among the various types of variables, which a bar diagram does not provide.

After selecting the limited factorial model, we verified its characteristics in view of the absolute contribution of the categories to each factor and the relative contributions of each category to the building of each axis. The suitability of the factorial model is checked by examining the variances shared by the chosen factorial axes, for each variable, and their correlation matrix.

The comparison among the categories comprising factorial planes and axes shows different features associated with the risk and injury variables, which we shall interpret at a later stage.

We have also corrected the active symmetries of the factor axes, which have appeared at some point in the process of the systematic dimensional reduction. These active symmetries alter the position of the group set without affecting their relative positions (Real 2001). Therefore, by means of an adequate homographic transformation, it returned to their original position without altering the features of the obtained initial settlement.

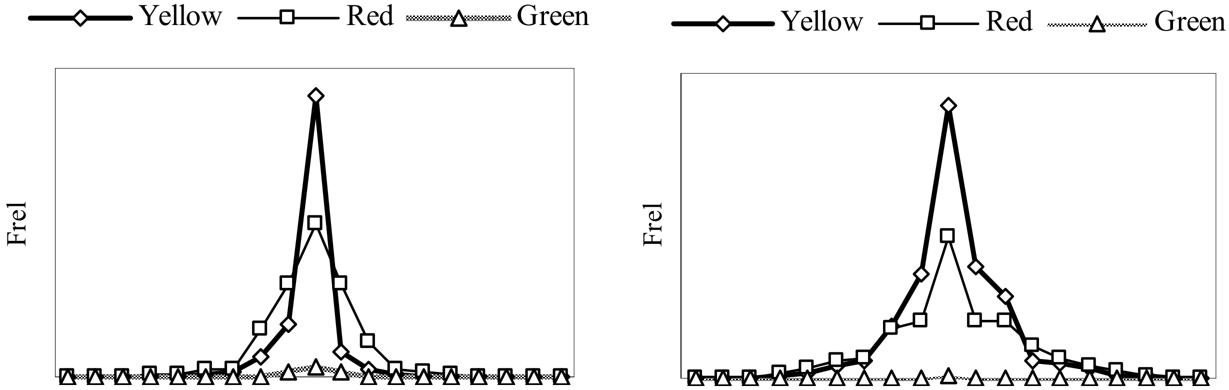
To quantify the affinity relationships among different modes, the distances between two points are calculated by using the Minkowski distance widespread on the factor coordinates. This is necessary because the graphic three-dimension representation is rather complex and, in two dimensions, errors can easily be made due to a deformation of the distances among modes that is caused by the orthogonal projection method in use. Therefore, it is possible to analytically prioritize the

TABLE I
Occupational accidents in Spain (average year).

Total	Injury	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19
Risk	683615	63688	14766	184772	62701	995	10282	2569	118680	33253	127549	35781	5107	14591	1024	299	237	586	185	6550
R1	59243	11496	2013	21362	1012	45	1518	44	3441	3114	13174	155	32	137	15	17	23	8	2	1635
R2	67195	9431	2207	29058	1184	46	1205	44	4524	3562	14937	126	27	204	10	21	3	5	2	599
R3	9478	1612	197	1112	179	7	278	24	1614	690	3404	96	18	48	2	3	6	1	0	187
R4	44618	9018	819	4756	293	34	714	128	8969	3346	14908	488	95	804	8	10	3	4	2	219
R5	4863	724	83	500	41	2	129	12	944	418	1649	205	21	54	1	1	1	1	0	77
R6	39361	2068	544	27370	320	13	268	28	5185	791	2505	65	15	98	3	10	1	3	1	73
R7	28452	2407	665	4727	226	16	815	47	7499	2731	8686	325	62	99	6	10	2	2	3	124
R8	16512	1468	349	2414	111	11	420	112	5378	1428	4289	209	38	89	2	6	1	1	1	185
R9	131419	11837	1632	13566	693	66	1882	503	51294	10141	37011	1552	334	361	18	35	3	10	6	475
R10	42887	221	42	327	76	7	133	27	5357	677	768	31064	2549	1503	18	0	7	35	14	62
R11	45490	7778	561	3273	112	22	414	1510	13938	2582	14651	123	23	221	4	22	10	3	2	241
R12	18776	2896	447	3889	287	16	657	68	1478	1352	5420	79	20	96	1	11	4	5	3	2047
R13	153039	2236	5097	71552	58099	706	1613	0	6286	2026	5075	0	0	0	0	50	12	0	0	287
R14	6677	0	0	0	0	0	43	2	458	0	0	0	0	6064	0	75	3	0	9	23
R15	2181	16	8	52	14	1	7	3	84	12	32	102	380	935	4	2	2	502	12	13
R16	7178	0	0	0	0	0	33	2	1074	78	0	1098	1054	2783	867	0	134	0	14	41
R17	704	0	0	0	0	0	6	1	84	10	0	0	402	82	1	3	0	0	112	3
R18	1314	32	4	42	5	0	25	6	115	27	51	45	21	851	19	6	15	6	1	43
R19	4228	448	98	772	49	3	122	8	958	268	989	49	16	162	45	17	7	0	1	216

Average year value = 683612; to round off value = 683615; total error to a round number = 3.

Figure 1 Figure 2



Contingency analysis
Multinomial distribution of 3 risks
(R15, green; R6, yellow; R4, red)

Contingency analysis
Multinomial distribution of 3 injuries
(I17, green; I3, yellow; I10, red)

Figs. 1 and 2 – Distribution of types of risks and injuries.

injury group concerning a specific risk. It means that, it is possible to define an order based on the proximity of the injuries with regard to the above-mentioned risk, or vice versa, to define an order based on the proximity of the injury compared to this risk and vice versa. It is possible to define an order based on the proximity of the risk compared to an injury.

RESULTS

The factors or dimensions obtained by the correspondence analysis for the variables under consideration are shown in Table II, along with the percentage of total variance that explains each factor.

The absence of the trivial solution ($\lambda_1 = 1$) shows that the analysis has been carried out on the centers of gravity of rows and columns. As the rows are quasi-barycenter of the columns and vice versa, they allow simultaneous graphic representation.

We selected the first three factors as a limited model (Table II), representing 72.9% of the total variance of the risk and injury variables. Therefore, we fulfilled Hair's criterion (Hair et al. 1999), which recommends that all the dimensions with inertia greater than 0.2 are selected.

Figure 3 shows the spatial disposition of the three axes or dimensions chosen as a solution. Each factorial axis is composed of a linear combination of the categories belonging to two different groups. Therefore, some of categories in the green group characterize Dim1's positive side, and some of the categories in the yellow group characterize its negative side. Dim2's positive side is characterized by the rest of the categories in the green group, and in its negative side by the categories in the red group. The red group characterizes Dim3's positive side, while its negative side is characterized by category R_{13}, which belongs to the yellow group. Therefore, we defined three semi-planes corresponding to each of the mentioned groups (Table III).

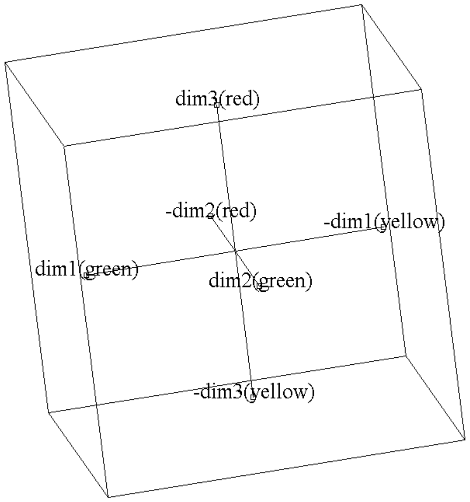


Fig. 3 – Characteristics of each factor according to the component categories.

TABLE II
Criterion for reduced factor model.

Dimension ($\psi i, \varphi i, \lambda i$)	Latent roots (μi)	Inertia (λi)	Variance explained	Cumulative proportion
1	0.85341	0.72831	0.313	0.313
2	0.71864	0.51644	0.222	0.535
3	0.67267	0.45248	0.194	0.729
4	0.44163	0.19504	0.084	0.813
5	0.38292	0.14662	0.063	0.876
6	0.32596	0.10625	0.046	0.922
7	0.30171	0.09103	0.039	0.961
8	0.22744	0.05173	0.022	0.983
9	0.14818	0.02196	0.009	0.992
10	0.11712	0.01372	0.006	0.998
11	0.04916	0.00242	0.001	0.999
12	0.02820	0.00080	0.000	1.000
13	0.02111	0.00045	0.000	1.000
14	0.01414	0.00020	0.000	1.000
15	0.01320	0.00017	0.000	1.000
16	0.00836	0.00007	0.000	1.000
17	0.00077	0.00000	0.000	1.000
18	0.00001	0.00000	0.000	1.000
Total		2.32767	1.000	1.000

ψi = row scores; φi = column scores; λi = eigenvalues.

TABLE III
Risks and injuries components of axes and planes.

	Dim1	Dim2	Dim3
Hyperplane (Dim1, Dim2)	R_{10}, R_{16}, R_{17} I_{11}, I_{12}, I_{18}	R_{14}, R_{15}, R_{18} $I_{13}, I_{14}, I_{15}, I_{16}$ I_{17}	
Hyperplane (Dim1, Dim3)	R_1, R_2, R_6 I_2		R_{13} I_1, I_4, I_5, I_6
Hyperplane (Dim2, Dim3)		R_{19}	R_3, R_4, R_5, R_7, R_8 $R_9, R_{11}, R_{12}, R_{19}$ $I_1, I_7, I_8, I_9, I_{10}$ I_{19}

This three-group solution (Table III) is well characterized by each of the three factorial hyperplanes: group 1 (green), hyperplane (Dim1, Dim2); group 2 (yellow), hyperplane (Dim1, Dim3); and group 3 (red), hyperplane (Dim2, Dim3). The factorial axes show a mixture of categories: Dim1 (green and yellow categories), Dim2 (green and red categories), and Dim3 (yellow and red categories). The fact that Dim2 represents few categories in the green group is a consequence of the low-shared variance of these categories with the rest of their group. This difference also classifies the various characteristics into two sub-groups of green categories.

Tables IV and V give the factor scores obtained for each of the studied variables, which are calculated for the first three dimensions and are sufficient to project the three stated groups. This three-dimensional approach is the one that defines the correct affinity relation-

TABLES IV and V

Factor Scores

Table IV – Risk variables.

Risk	Dim. 1	Dim. 2	Dim. 3
R1	–0.318	0.012	0.200
R2	–0.335	0.013	0.080
R3	–0.172	–0.009	0.672
R4	–0.108	0.065	0.731
R5	–0.003	–0.062	0.675
R6	–0.385	0.012	–0.345
R7	–0.153	–0.034	0.627
R8	–0.119	–0.028	0.670
R9	–0.101	–0.051	0.789
R10	2.906	–1.764	–0.491
R11	–0.148	–0.005	0.893
R12	–0.244	0.032	0.428
R13	–0.543	0.014	–1.320
R14	3.035	5.934	–0.417
R15	3.078	4.372	–0.625
R16	2.797	2.762	–0.327
R17	2.883	1.387	–0.395
R18	2.382	4.271	–0.229
R19	0.005	0.272	0.480

Table V – Injury variables.

Injury	Dim. 1	Dim. 2	Dim. 3
I1	–0.237	0.002	0.649
I2	–0.382	0.007	–0.289
I3	–0.439	0.011	–0.583
I4	–0.605	0.018	–1.782
I5	–0.491	0.003	–1.193
I6	–0.213	0.036	0.273
I7	–0.115	–0.008	1.150
I8	0.021	–0.057	0.731
I9	–0.144	–0.043	0.650
I10	–0.199	–0.018	0.744
I11	3.056	–1.190	–0.556
I12	2.900	0.194	–0.446
I13	2.838	4.666	–0.366
I14	2.888	3.363	–0.354
I15	0.784	2.271	–0.092
I16	2.108	2.646	–0.245
I17	3.303	5.125	–0.793
I18	2.947	2.103	–0.438
I19	–0.191	0.109	0.481

ship among the analyzed variables. We must be careful in interpreting the above-mentioned relationships using the two-dimensional projections from the decomposition of the bucket solution, due to the distortions that the plane projection imposes on the results space.

The absolute contribution of a variable to a dimension indicates the percentage of inertia (variance) of this dimension attributable to the above-mentioned variable. Tables VI and VII show the variables that are most important or best characterize the chosen dimensions. Therefore, we observed that, for the risk variables, 72.6% of the inertia of Dim1 are due to 62.1% from R_{10} (projection of fragments or particles), and 10.5% from R_{14} (exposure to heat contacts). For Dim2, 86.2% of their inertia are distributed among three variables: 47.9% from R_{14} (exposure to heat contacts), 27.2% from R_{10} (projection of fragments or particles), and 11.1% from R_{16} (exposure to chemical contacts). In the case of Dim3, both variables represent 75.8% of their inertia: 58% from R_{13} (overexertion) and 17.8% from R_9 (bruises, contusions and cuts by objects or tools).

For the injury variables, 77.4% of the inertia of Dim1 are distributed between two variables: 57.3% from I_{11} (objects in the eyes), and 20.1% from I_{13} (burns). For Dim2, 93.7% of their inertia are distributed among 28.9% from I_{11} (objects in the eyes) and 64.8% from I_{13} (burns). For Dim3, 82.3% of their inertia are distributed among 43.3% from I_4 (back pain), 15.4% from I_{10} (bruises, contusions and crushing), 13.8% from I_8 (other injuries), and 13.7% from I_3 (twists, sprains and strains).

Figures 4 and 5 represent the risk forms (R_{10} for Dim1, R_{14} for Dim2 and R_{13} and R_9 for Dim3) and injury forms (I_{11} for Dim1, I_{13} for Dim2, and I_4 and I_{10} for Dim3) respectively, which most contribute to the formation of each axis where the centroids of the groups are located. The stated forms have projected orthogonally to three planes formed by three axes solution.

Considering the risk and injury variables together, and for Dim1, the relationship of R_{10} with I_{11} (the variables that most contribute to the inertia) is noticed, which perfectly explains the qualitative meaning or affinity of this risk-injury pair. In the same way, the rela-

TABLES VI and VII

Absolute contributions

Table VI – Contribution of row points to the inertia of each dimension.

Risk	Marginal	Dim. 1	Dim. 2	Dim. 3
R1	0.087	0.010	0.000	0.005
R2	0.098	0.013	0.000	0.001
R3	0.014	0.000	0.000	0.009
R4	0.065	0.001	0.000	0.052
R5	0.007	0.000	0.000	0.005
R6	0.058	0.010	0.000	0.010
R7	0.042	0.001	0.000	0.024
R8	0.024	0.000	0.000	0.016
R9	0.192	0.002	0.001	0.178
R10	0.063	0.621	0.272	0.023
R11	0.067	0.002	0.000	0.079
R12	0.027	0.002	0.000	0.009
R13	0.224	0.077	0.000	0.580
R14	0.010	0.105	0.479	0.003
R15	0.003	0.035	0.085	0.002
R16	0.011	0.096	0.111	0.002
R17	0.001	0.010	0.003	0.000
R18	0.002	0.013	0.049	0.000
R19	0.006	0.000	0.001	0.002
Total		1.000	1.000	1.000

Table VII – Contribution of column points to the inertia of each dimension.

Injury	Marginal	Dim. 1	Dim. 2	Dim. 3
I1	0.093	0.006	0.000	0.058
I2	0.022	0.004	0.000	0.003
I3	0.270	0.061	0.000	0.137
I4	0.092	0.039	0.000	0.433
I5	0.001	0.000	0.000	0.003
I6	0.015	0.001	0.000	0.002
I7	0.004	0.000	0.000	0.007
I8	0.174	0.000	0.001	0.138
I9	0.049	0.001	0.000	0.031
I10	0.187	0.009	0.000	0.154
I11	0.052	0.573	0.289	0.024
I12	0.007	0.074	0.000	0.002
I13	0.021	0.201	0.647	0.004
I14	0.001	0.015	0.024	0.000
I15	0.000	0.000	0.003	0.000
I16	0.000	0.002	0.003	0.000
I17	0.001	0.011	0.031	0.001
I18	0.000	0.003	0.002	0.000
I19	0.010	0.000	0.000	0.003
Total		1.000	1.000	1.000

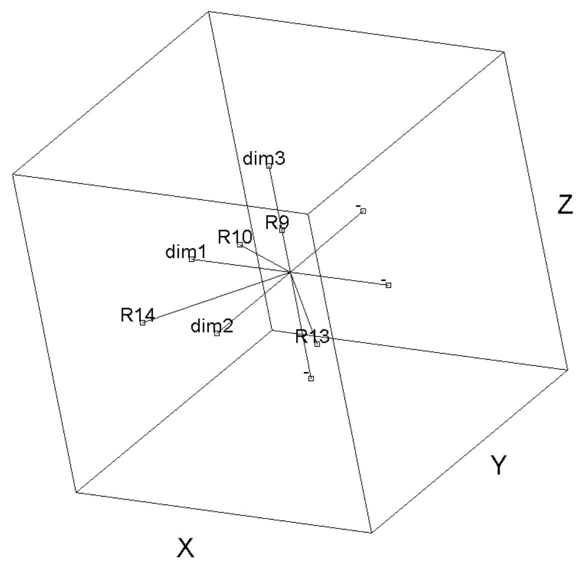


Fig. 4 – Risks {R_{14}, R_{10}, R_{13} (R_9)} with a major contribution to the inertia of each dimension.

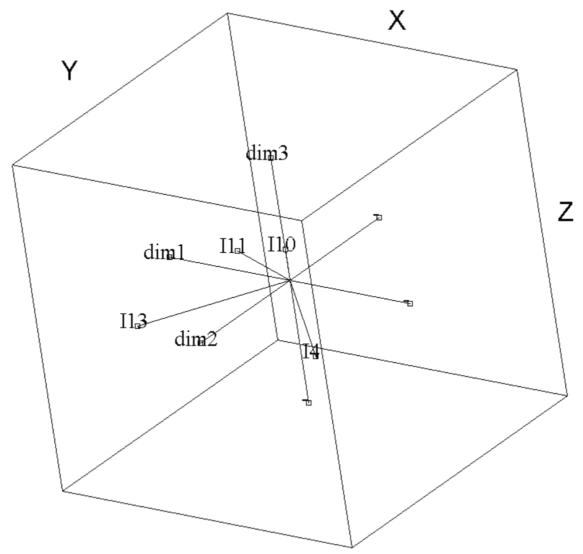


Fig. 5 – Injuries {I_{11}, I_{13}, I_4 (I_{10})} with a major contribution to the inertia of each dimension.

TABLES VIII and IX

Relative contributions

Table VIII – Contribution of dimensions to inertia of each row point.

Risk	Marginal	Dim. 1	Dim. 2	Dim. 3	Total
R1	0.087	0.211	0.000	0.066	0.277
R2	0.098	0.285	0.000	0.013	0.298
R3	0.014	0.055	0.000	0.664	0.720
R4	0.065	0.021	0.006	0.743	0.769
R5	0.007	0.000	0.007	0.779	0.786
R6	0.058	0.132	0.000	0.083	0.215
R7	0.042	0.056	0.002	0.743	0.802
R8	0.024	0.031	0.002	0.784	0.817
R9	0.192	0.015	0.003	0.720	0.738
R10	0.063	0.747	0.232	0.017	0.995
R11	0.067	0.023	0.000	0.664	0.688
R12	0.027	0.037	0.001	0.115	0.153
R13	0.224	0.170	0.000	0.794	0.965
R14	0.010	0.207	0.666	0.003	0.876
R15	0.003	0.110	0.186	0.004	0.300
R16	0.011	0.329	0.270	0.004	0.603
R17	0.001	0.052	0.010	0.001	0.063
R18	0.002	0.248	0.672	0.002	0.922
R19	0.006	0.000	0.115	0.333	0.448

Table IX – Contribution of dimensions to inertia of each column point.

Injury	Marginal	Dim. 1	Dim. 2	Dim. 3	Total
I1	0.093	0.089	0.000	0.527	0.616
I2	0.022	0.481	0.000	0.217	0.698
I3	0.270	0.301	0.000	0.420	0.721
I4	0.092	0.110	0.000	0.751	0.861
I5	0.001	0.150	0.000	0.698	0.848
I6	0.015	0.148	0.004	0.193	0.344
I7	0.004	0.002	0.000	0.195	0.197
I8	0.174	0.001	0.004	0.600	0.604
I9	0.049	0.051	0.004	0.826	0.881
I10	0.187	0.079	0.001	0.864	0.943
I11	0.052	0.716	0.256	0.019	0.991
I12	0.007	0.484	0.002	0.009	0.495
I13	0.021	0.293	0.666	0.004	0.963
I14	0.001	0.105	0.120	0.001	0.226
I15	0.000	0.077	0.546	0.001	0.624
I16	0.000	0.119	0.158	0.001	0.278
I17	0.001	0.041	0.082	0.002	0.125
I18	0.000	0.021	0.009	0.000	0.030
I19	0.010	0.008	0.002	0.041	0.052

tionships are verified among many studied variables: R_{14} with I_{13} (Dim2), R_{13} with I_4, and R_9 with I_{10} (Dim3). The variables that most influenced the inertia of the dimension are also those that are near to the centroid.

The relative contribution of a dimension to a variable, Tables VIII and IX, represents the correlation's measure between the dimension and the variable. This indicates the proportion of the inertia of the variable explained by the dimension. The amount of relative contributions of a variable is equivalent to the concept of its shared variance (communality) used in the classic factor analysis.

In the case of risk variables (Table VIII), the worst represented or those with the worst reconstitution quality as three chosen dimensions are R_1 (27.7%), R_2 (29.8%), R_6 (21.5%), R_{12} (15.3%), R_{15} (30.0%), and R_{17} (6.3%).

In the case of injury variables (Table IX), the worst represented considering three chosen dimensions are I_6 (34.4%), I_7 (19.7%), I_{14} (22.6%), I_{16} (27.8%), I_{17} (12.5%), I_{18} (3.0%), and I_{19} (5.2%).

Therefore, Dim3 is the one that best represents the diversity of all variables that compose the analyzed table.

This situation shows the discontinuity that appears in the frequencies of the initial contingency table, one of high and one of low frequencies. The sub-table of high frequencies is largely characterized by Dim3, while the low frequencies are represented by the other two dimensions.

With regard to the data, the shared variance among the chosen factorial axes has been analyzed, for each variable, as well as the correlations matrix among them. Both shared variances and obtained correlations indicate the independence among the factorial axes and, therefore, a suitable representation of the information by the model, that is, the stability of the adopted solution.

Figures 6 and 7 present the previously achieved results. The aim is to verify the scattering of groups.

Table X shows the relationships of affinity between the risk and injury vectors calculated as Minkowski distances, forming a decision criterion of great interest in their forecasting and prioritizing.

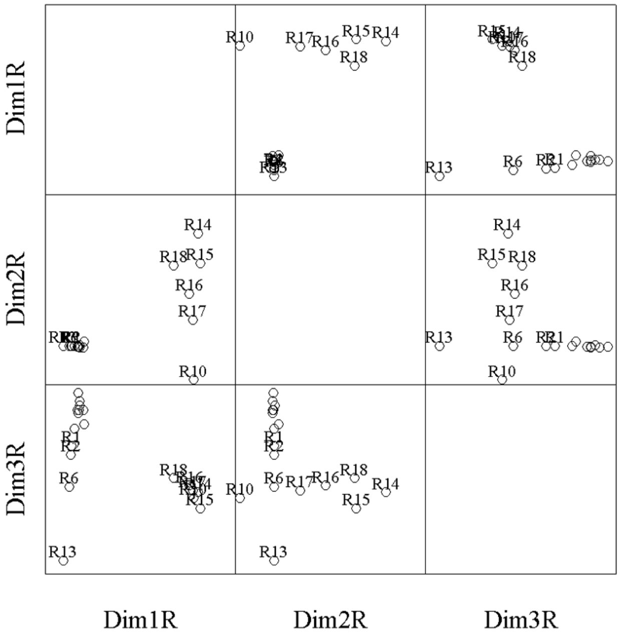


Fig. 6 – Factor model: groups of risks for three axes. (Red group without label; points projected on coordinate planes).

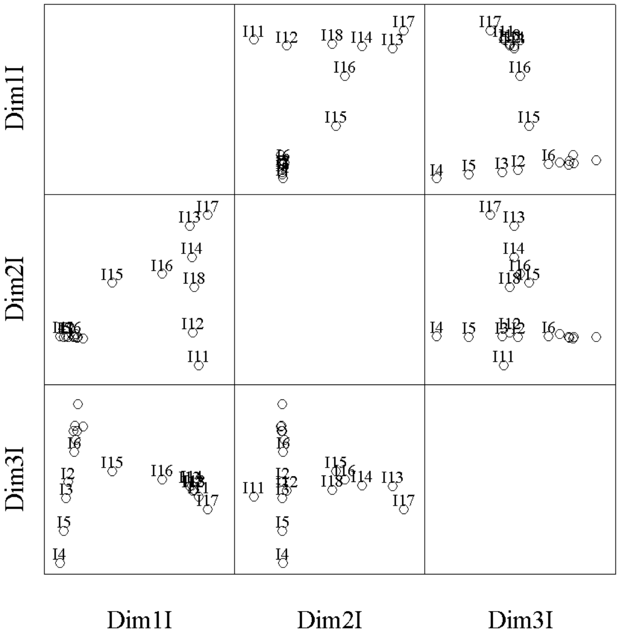


Fig. 7 – Factor model: groups of injuries for three axes. (Red group without label; points projected on coordinate planes).

DISCUSSION

The presented methodology overcomes certain limitations imposed by classical analytical methods regarding accident rates: "free risk assessment methods", that use tables valued on a qualitative or quantative ordinal scale, but with major limitations that impose the direct and subjective assignment of risk values and "logical methods" based on the analysis of probability trees (event and fault trees) where the majority of starting probabilities are usually estimated and not calculated on observations.

This methodology allows the calculation of probabilities associated with the diverse nature of occupational accident rates, being the basis for an in-depth ana-

TABLE X
Affinity relationships (proximities) between risk and injury vectors.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19
I1	0.456	0.577	0.069	0.165	0.244	1.005	0.094	0.123	0.202	3.781	0.259	0.223	1.992	6.857	5.631	4.261	3.569	5.084	0.400
I2	0.493	0.372	0.983	1.057	1.038	0.056	0.945	0.995	1.115	3.740	1.205	0.734	1.043	6.842	5.580	4.206	3.546	5.081	0.900
I3	0.792	0.671	1.283	1.356	1.333	0.244	1.244	1.293	1.414	3.787	1.504	1.033	0.744	6.868	5.602	4.255	3.600	5.121	1.181
I4	2.002	1.881	2.492	2.562	2.530	1.453	2.451	2.500	2.620	4.143	2.713	2.242	0.466	7.079	5.819	4.606	3.995	5.424	2.356
I5	1.403	1.282	1.892	1.962	1.931	0.854	1.851	1.900	2.020	3.892	2.114	1.643	0.137	6.943	5.670	4.378	3.733	5.234	1.765
I6	0.130	0.229	0.403	0.470	0.464	0.641	0.365	0.413	0.535	3.681	0.624	0.155	1.627	6.768	5.517	4.105	3.443	4.992	0.382
I7	0.971	1.092	0.481	0.425	0.491	1.519	0.525	0.480	0.363	3.860	0.259	0.731	2.506	6.905	5.703	4.281	3.649	5.142	0.736
I8	0.633	0.745	0.207	0.177	0.061	1.152	0.204	0.155	0.135	3.568	0.239	0.399	2.128	6.804	5.549	4.095	3.397	5.022	0.414
I9	0.485	0.603	0.049	0.139	0.144	1.025	0.026	0.035	0.145	3.683	0.246	0.247	2.010	6.853	5.612	4.180	3.507	5.075	0.387
I10	0.557	0.678	0.077	0.123	0.212	1.105	0.126	0.109	0.112	3.770	0.158	0.320	2.092	6.872	5.646	4.225	3.573	5.099	0.442
I11	3.660	3.653	3.650	3.639	3.485	3.651	3.610	3.596	3.615	0.596	3.710	3.635	3.871	7.125	5.562	3.967	2.587	5.512	3.538
I12	3.287	3.282	3.275	3.232	3.122	3.291	3.244	3.226	3.254	1.958	3.335	3.248	3.556	5.741	4.185	2.572	1.194	4.115	3.040
I13	5.651	5.649	5.656	5.572	5.613	5.661	5.658	5.643	5.676	6.431	5.685	5.610	5.829	1.284	0.459	1.904	3.279	0.618	5.296
I14	4.670	4.668	4.667	4.585	4.598	4.684	4.663	4.646	4.679	5.128	4.702	4.625	4.890	2.576	1.061	0.608	1.976	1.047	4.308
I15	2.530	2.525	2.587	2.517	2.578	2.556	2.590	2.584	2.636	4.576	2.649	2.509	2.891	4.311	3.156	2.085	2.297	2.563	2.220
I16	3.608	3.606	3.617	3.539	3.554	3.628	3.613	3.598	3.636	4.488	3.662	3.567	3.887	3.420	2.016	0.703	1.486	1.648	3.253
I17	6.343	6.334	6.370	2.289	6.323	6.320	6.369	6.356	6.393	6.907	6.408	6.314	6.418	0.931	0.803	2.461	3.782	1.376	6.004
I18	3.929	3.925	3.926	3.854	3.824	3.934	3.912	3.894	3.928	3.867	3.974	3.885	4.162	3.832	2.280	0.684	0.720	2.250	3.584
I19	0.322	0.436	0.225	0.267	0.319	0.854	0.207	0.244	0.358	3.747	0.429	0.099	1.837	6.718	5.484	4.076	3.442	4.944	0.254

lysis of the observed frequencies and greatly exceeding the analytical expectations of the currently used methods as indicated in the above paragraph.

The correspondence analysis is the core or central body of this new methodology, but its full development exceeds the parameters of this article. A basic concept that it provides is that of a "population pattern of accident rates" named acsom, which can include various studies aimed at characterizing (the study of masses and potentials), comparing (the study of company-pattern deviations) and controlling (preventive action plans) corporate accident rates, mainly regarding the frequency of their occurrence and their seriousness.

As indicated by Schroeder-Frechette (1999), the multi-dimensional approach to risk must take into account the following ethical problems: (1) who defines the risk and how it should be defined, (2) who evaluated the risk and in accordance with what rules, and (3) under what conditions is ethically acceptable to impose risks upon the society.

The problem of occupational accidents is a restricted variant on the problem of risk, as defined by Giddens (1994) and Beck (1999), appearing in their manifestation at least as a reflection in the physical-natural world.

Regardless the existence of human beings, the risk of accidents will continue to exist as a natural phenomenon that may happen, as in fact they do, to any other biological species. The frequency at which they occur (the accident rate) is increased by social-economic activity, and their control is ethically obligatory as they cause injuries of varying intensity (seriousness) to the health of individuals.

The correspondence model, or joint probability model, of accident rates in the Spanish workforce reproduces some risk-injury groups similar to those obtained through other analyses based on the study of rows and columns: principal components, multidimensional scaling and hierarchical clustering analysis. This is what best establishes the relationships between risk and injury, being confirmed as the most suitable analytical method to treat the exposed problems under the proposals already outlined.

As for the groups, the group-1 or green includes all those risk and injury variables related to technological problems of recent historical appearance (the industrial revolution) and related to scientific and technical development (Baram 2009, Rasmussen 1997). The group-2 or yellow group contains all those risk and injury variables related to evolutionary biomechanical problems (Nachreiner et al. 2006). The group-3 or red group contains all those risk and injury variables related to technical-cultural problems (Guldenmund 2000) and to the evolution of their activity.

The groupings also match the results obtained by Douglas and Wildavsky (1982) by indicating the lack of differences between the hazards that were posed in early history (red and yellow groups) and those from developed civilizations (red, yellow and green groups), excepting in the type of cultural perception and the way in which a civilization has organized itself into a global society.

RISK AND INJURY VARIABLES OF THE FIRST FACTOR

The variables of the first factor contribute to the formation of the positive side of dimension 1, the categories of the green group $\{R_{10}, R_{16}, R_{17}\}$, and the formation of the negative side of the yellow group $\{R_1, R_2, R_6\}$. This dimension is associated to projections of fragments at the macroscopic or microscopic scale, to the exposure to solid, liquid or gas chemical substances, to radiation or to exposure at a subatomic scale (green group), with the accident rate for anomalies of gravitational interaction, fall of persons and treading on objects (yellow group). This dimension can be interpreted as the physical process "projections" from the environment on the individual (green group) or from the individual on the environment (yellow group).

RISK AND INJURY VARIABLES OF THE SECOND FACTOR

Axis 2 is formed by the linear combination of variables of the green group $\{R_{14}, R_{15}, R_{18}\}$ whose common factor is thermal effects and the resultant injury mostly being burns (trauma-type, thermal-type), which contributes to the positive side. Similarly, on the other side of the variable, the R_{19} from the red group, in which the fewest injuries are of trauma-type generated by the interaction with living beings, including falls, bruises, strokes, blows, shocks, bites, stings, etc., is the main contribution to the negative side of the axis.

RISK AND INJURY VARIABLES OF THE THIRD FACTOR

Axis 3 is formed by the yellow group, with R_{13} forming its negative side, and its positive side is formed by the variables of the red group $\{R_3, R_4, R_5\}$ that represent "fall of objects", a by $\{R_7, R_8, R_9\}$ that represent bruises, blows and collisions against or by objects. R_{11} represents the cases of being caught by objects, and R_{12} represents the cases related to mobile machinery and traffic in the workplace (excluding accidents while travelling). Axis 3 represents the trauma-type injuries caused by external agents to the individual, or by the individual himself.

FACTORS COMPOSITION: FACTORIAL PLANES

The combination of the factorial axes defines three factorial hyper-planes, which represents the risk and injury groups. Therefore, axes 1 and 2 define the green group (environmental risk and injury), axes 1 and 3 define the yellow group (risks associated with individual and muscle-skeletal injuries), and finally axes 2 and 3 define the red group (individual mixed risk and trauma-type injuries).

GREEN GROUP: PLANE (DIM1, DIM2)

The industrial accident identifies this group as risks associated with the work environment and injuries caused by the environment.

The green group is characterized by the low occurrence of the frequencies of its component variables, the temporal instability of their relative frequencies and the highest accumulation of mass in two or three injury variables. Figures 1 and 2 represent the multinomial distributions corresponding to a risk (R_{15}) and an injury (I_{17}), respectively, from the green group. Individuals are presented as a passive element in the individual interaction environment, without the ability to respond to an accident.

YELLOW GROUP: PLANE (DIM1, DIM3)

The yellow group is characterized by the high occurrence of risk, the temporal stability of its relative frequencies, and the highest accumulation of mass in one or two injury variables (heterogeneity in the distribution). Figures 1 and 2 represent the multinomial distributions corresponding to a risk (R_6) and an injury (I_3), respectively, from the yellow group. Individuals are presented as an (dynamic) active element in the individual-environment interaction, and responsive to the accident. The environment will be a (static) passive element.

RED GROUP: PLANE (DIM2, DIM3)

The red group is characterized by the high occurrence of the risk, the temporal stability of its relative frequencies, and the distribution of the principal mass in 5 or more categories (greater homogeneity in the distribution). Figures 1 and 2 represent the discrete multinomial distributions corresponding to a risk (R_4) and an injury (I_{10}), respectively, from the red group. Both the individual and environment can be active elements in the interaction.

CONCLUSIONS

The presented risk-injury correspondence model results in three groups of risks and injuries. The advantage over other factorial models stems from the joint treatment of risks and injuries, thereby obtaining groupings composed of the variables. These three groups, called green (technological/environmental), yellow (biological/evolutionary) and red (technical/cultural) groups, have been verified as a necessary and sufficient condition for the abbreviated representation of occupational accident rates.

These risk and injury groupings define a pattern that we called "accident soma" or acsom-G, which should be understood as a global model that represents the balancing conditions of occupational accidents in a population, enabling multiple specific analyses of companies to be carried out.

Based on the presented result, new possibilities are opened for the development of applications focused on the automatic analysis, interpretation and management of occupational accidents, thereby minimizing uncertainty and improving the objectivity not offered by current methods.

RESUMO

Apresentamos aqui um modelo generalizado para o diagnóstico e predição de acidentes na classe de trabalhadores da Espanha. Baseados em dados sobre a frequência de acidentes em todas as companhias da Espanha em 11 anos (7.519.732 acidentes), nós os classificamos em uma nova tabela de contingência risco-injúria (19×19). Através de uma análise por correspondência obtivemos uma estrutura composta por 3 eixos cuja combinação identifica 3 grupos separados de risco e injúria, que nós usamos como um perfil geral na Espanha. As mais prováveis ou frequentes relações entre risco e injúrias identificadas nesse perfil facilitaram o processo de decisão nas companhias em um estágio inicial de apreciação do risco. Cada grupo de risco-injúria tem suas próprias características que são compreensíveis dentro do conteúdo fenomenológico do acidente. As principais vantagens desse modelo são a sua aplicação potencial em qualquer outro País e a possibilidade de comparar resultados de diferentes países. Um fator limitante, contudo, é a necessidade de se usar um padrão comum de classificação para riscos e injúrias afim de facilitar a comparação, um padrão que não existe hoje. O modelo tem como alvo administrar acidentes ligados ao trabalho automaticamente em qualquer nível.

Palavras-chave: modelo de correspondência, análise de contingência, risco, injúria, acidentes ocupacionais.

REFERENCES

AGUILERA AM. 2001. Tablas de contingencia bidimensionales. Madrid: La Muralla, Col. Cuadernos de Estadística 15: 23–36.

AGUILERA AM. 2006. Modelización de tablas de contingencia multidimensionales. Madrid: La Muralla, Col. Cuadernos de Estadística 33: 25–28.

AMENDOLA A. 2002. Recent paradigms for risk informed decision making. Safety Sci 40: 17–30.

BARAM M. 2009. Globalization and workplace hazards in developing nations. Safety Sci 47: 756–766.

BARIL R, BERTHELETTE D AND MASSICOTTE P. 2003. Early return to work of injured workers: multidimensional patterns of individual and organizational factors. Safety Sci 41: 277–300.

BECK U. 1999. World risk society. Cambridge: Polity Press.

BENZÉCRI J. 1992. Correspondence analysis handbook. New York: Marcel Decker.

CONTE JC, CANO F, GARCIA AI AND RUBIO E. 2008. Interpretación de las relaciones intragrupales de riesgos y lesiones mediante análisis cluster jerárquico. Rev Mat Teor Aplic 15: 175–186.

CONTE JC, MARCOS G, GARCIA AI AND RUBIO E. 2007. Análisis del problema empírico de identificación del riesgo. Cuad Bioest Aplic Infor 17: 12–25.

DOUGLAS M AND WILDAVSKY A. 1982. Risk and culture. An essay on the selection of technical and environmental dangers. Berkeley: University of California Press.

FINE W. 1973. Mathematical evaluations for controlling hazards. Georgia: Academic Press, Selected Reading in Safety, p. 1–22.

FRIJTERS A AND SWUSTE P. 2008. Safety assessment in design and preparation phase. Safety Sci 46: 272–281.

GARCIA AI, CONTE JC, RUBIO E AND PEREZ A. 2009. Accidente laboral. ACSOM una nueva orientación para la gestión automática del riesgo. An Sist Sanit Navar 32: 23–34.

GIDDENS A. 1994. Les conséquences de la modernité. Paris: L'Harmattan.

GREENACRE M AND BLASIUS J. 1994. Correspondence analysis in social science. Recent developments and applications. San Diego: Academic Press, p. 141–161.

GULDENMUND FW. 2000. The nature of safety culture: a review of theory and research. Safety Sci 34: 215–257.

HAIR JF, ANDERSON RE, TATHAM RL AND BLACK WC. 1999. Análisis Multivariante. Madrid: Prentice Hall Iberia SRL, 5th ed., p. 575–578.

HAMMER W. 1994. Unfallgefährdung und verhütung beim Gehen, Laufen, Tragen, Schieben und Ziehen im landwirtschaftlichen Betrieb. Safety Sci 17: 117–143.

HAND D, MANNILA H AND SMITH P. 2001. Principles of data mining. London: Bradford Book, MIT Press.

JOARISTI L AND LIZASOAIN L. 1999. Análisis de correspondencias. Madrid: La Muralla, Col. Cuadernos de Estadística 5: 35–52.

KJELLÉN U AND SKLET S. 1995. Integrating analyses of the risk of occupational accidents into the design process. Part I: A review of types of acceptance criteria and risk analysis methods. Safety Sci 18: 215–227.

KÖRVERS P AND SONNEMANS P. 2008. Accidents: A discrepancy between indicators and facts! Safety Sci 46: 1067–1077.

LAFLAMME L, DÖÖS M AND BACKSTRÖM T. 1991. Identifying accidents patterns using the FAC and HAC: their application to accidents at the engine workshops of an automobile and truck factory. Safety Sci 14: 13–33.

LEVESON N. 2004. A new accident model for engineering safer systems. Safety Sci 42: 237–270.

NACHREINER F, NICKEL P AND MEYER I. 2006. Human factors in process control systems: The design of human machine interfaces. Safety Sci 44: 5–26.

NICHOLSON A. 1998. Analysis of spatial distributions of accidents. Safety Sci 31: 71–91.

RASMUSSEN J. 1997. Risk management in a dynamic society: a modelling problem. Safety Sci 27: 183–213.

REAL JE. 2001. Escalamiento multidimensional. Madrid: La Muralla, Col. Cuadernos de Estadística 14: 8–14.

ROUHIAINEN V. 1992. QUASA: A method for assessing the quality of safety analysis. Safety Sci 15: 155–172.

RUBIO E. 1983. Estadística para médicos: fundamentos. Universidad de Zaragoza, Cátedra de Bioestadística, Facultad de Medicina, p. 119–122.

SARI M, SELCUK S, KARPUZ C AND DUZGUN S. 2009. Stochastic modelling of accident risks associated with an underground coal mine in Turkey. Safety Sci 47: 78–87.

SCHROEDER-FRECHETTE K. 1999. Risk. In: CRAIG E AND FLORIDI L. Routledge Encyclopaedia of Philosophy. London: Routledge.

VAN DUIJNE F, VAN AKEN D AND SCHOUTEN G. 2008. Considerations in developing complete and quantified methods for risk assessment. Safety Sci 46: 245–254.

WILLIAMSON A, FEYER A AND CAIRNS DR. 1996. Industry differences in accident causation. Safety Sci 24: 1–12.