v. 8, n. 1, p. 7-23, 2019
ISSN 2237-9223

REVISTA BRASILEIRA DE CRIMINALÍSTICA

DOI: http://dx.doi.org/10.15260/rbc.v8i1.224

Analysis of the procedures used in the forensic evaluation of illegal substances: a statistical approach to interpreting reports

R.L. Yoshida [a], A.S.L Ferreira [b], J.A. Velho [b,c], B.S. De Martinis [a,b], V.B.P. Leite [d], A.T. Bruni [b,c*]

[a] *Departamento de Análises Clínicas, Toxicológicas e Bromatológicas, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, Ribeirão Preto (SP), Brasil*
[b] *Departamento de Química, Faculdade de Filosofia Ciências e Letras, Universidade de São Paulo, Ribeirão Preto (SP), Brasil*
[b] *Instituto Nacional de Ciência e Tecnologia Forense (INCT Forense), Ribeirão Preto (SP), Brasil*
[d] *Departamento de Física, Instituto de Biociências Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto (SP), Brasil*

**Endereço de e-mail para correspondência: aline.bruni@usp.br. Tel.: +55-16-33153689.*

Recebido em 27/09/2017; Revisado em 18/06/2018; Aceito em 03/04/2019

Resumo

Este estudo utilizou ferramentas estatísticas para avaliar laudos forenses sobre substâncias ilegais. Avaliamos variáveis quanto às características da análise e abordamos a metodologia empregada pelos peritos. Perguntas baseadas no que é necessário para esclarecer a lei foram formuladas. Analisamos 1008 documentos oficiais de diferentes jurisdições, divididos em 504 conjuntos compostos por um laudo preliminar e um laudo definitivo para cada caso. Os laudos foram examinados por uma equação empírica formulada para fornecer um parâmetro denominado "*Report Relevance*" (Relevância do Laudo), que teve por finalidade classificar cada documento de acordo com uma pontuação relacionada à quantidade de informação contida. A validação do método foi realizada por análise multivariada de dados: Análise de Componentes Principais (*Principal Component Analysis*, PCA), Análise de Agrupamentos Hierárquicos (*Hierarchical Cluster Analysis*, HCA), *Soft Independent Modeling of Class Analogy* (SIMCA) e Mínimos Quadrados Parciais (*Partial Least Squares*, PLS). A análise quantitativa mostrou que os documentos foram bem produzidos, com boa qualidade, uma vez que a Relevância do Laudo apresentou valores em torno de 0,74 ± 0,08 para aqueles provenientes da Polícia Estadual. Em comparação, os documentos provenientes da Polícia Federal obtiveram valores em torno de 0,87 ± 0,05. Fatores que podem explicar essas diferenças e as melhores pontuações para os laudos federais incluem maior investimento em tecnologia e treinamento de pessoal, e menor demanda de mão-de-obra e rotina. Para ambas as forças policiais, alguns aspectos poderiam ser melhorados, como imagens das evidências coletadas ou procedimentos analíticos laboratoriais. Finalmente, a metodologia neste estudo pode ser adaptada para ser usada em outros tipos de investigação forense.
Palavras-chave: Substâncias Ilícitas, Procedimentos Periciais, Análise Multivariada

Abstract

This study used statistical tools to evaluate forensic reports on illegal substances. We evaluated variables regarding the characteristics of the analysis and we addressed the methodology employed by the experts. Questions based on what is required to clarify the law were formulated. We have parsed 1008 official documents from different jurisdictions, divided into 504 sets comprised of a preliminary and a final report for each case. The reports were examined by an empirical equation formulated to provide a parameter called "Report Relevance", which intended to classify each report according to a score related to the amount of information it contained. The validation of the method was performed by multivariate data analysis: Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), Soft Independent Modeling of Class Analogy (SIMCA) and Partial Least Squares (PLS). Quantitative analysis showed that the expert documents were well produced, with good quality, since the Report Relevance showed values around 0.74 ± 0.08 for the reports from the State Police. By comparison, reports from the Federal Police obtained scores around 0.87 ± 0.05. Factors that might explain these differences and the better scores for the Federal reports include increased investment in technology and training of staff, and a lower labor demand and routine. For both police forces, some aspects could be improved, such as images of the collected evidence or laboratory analytical procedures. Finally, the methodology in this study can be adapted to be used in other kinds of forensic investigation.

Keywords: Illegal Substances, Expertise Procedures, Multivariate Analysis.

1. INTRODUCTION

Material evidence has increasingly gained focus and importance. When properly identified and analyzed, it is the best instrument, which allows the correct application of the law.

The forensic investigation process became more reliable with the development of scientific methods of analysis. The conclusion of an expert criminal investigation is usually recorded in a report, which is a document that must be prepared with technical rigor and substantiated by irrefutable elements. Such reports should be scientific and objective in order to be accepted in court.[1]

The quality of official reports is a delicate and little explored area. Internationally, attention has been given to results obtained from psychology and forensic psychiatry.[2–6] In such cases, it is important to assure quality, since different and even conflicting results may be obtained by different practitioners. Even techniques of forensic analysis considered to be determinant or practically free of flaws, such as DNA testing, may have subjective interpretations and possibly even be misleading.[7–9]

In Brazil, a few initiatives with this type of study have been undertaken.[10–13] A wide variation in the production of reports may be observed, and the desired information is not always present or easily interpreted. It is important to have a means of indicating the real quality of expert reports.

According to Brazilian Legislation (Law No. 11.343 of 2006) two types of report should be issued in cases of the seizure of illicit drugs[14]:

Preliminary report: a document produced for provisional characterization of the nature and amount of the substance. The analytical methods used in this step are simple, usually colorimetric. They are highly sensitive, but are not specific, and false-positive and false-negative results can often be found. This document is used as grounds for arrest in flagrant. If the scientific methodology presents non-specific results, unlawful arrests may occur.

a) Final report: produced by more robust methods, which are able to make an unequivocal identification of the drug. It should either confirm or correct the preliminary report.

The main objective of the present study was to evaluate expert reports generated in the seizure of drugs. Statistical methods were employed to interpret and quantify the technical information contained in official documents. The use of statistical tools in forensic science is of the utmost importance in judicial processes. Forensic scientists can assess and interpret the evidence, which often includes elements of uncertainty. They increasingly rely on statistical science in its various branches.[15–17]

2. MATERIAL AND METHODS

We analyzed 1008 reports, divided into 504 subgroups each containing a preliminary and final document about a particular case. Of these, 439 were from State jurisdictions and 65 came from Federal jurisdictions.

The main difference between the jurisdictions is that, at the State level, trafficking inside the territory of each state is investigated, whereas, in the Federal jurisdiction, drug traffic between different states of the federation is investigated, as well as the borders of the country.

Table 1. Variables studied in the analysis of preliminary reports.

V01	Was the nature of the substance indicated in the report?
V02	Was there some photo attached of the material seized?
V03	Was there some information about the amount (mass) of the substance?
V04	Was the substance physically described?
V05	Were packaging and wrappers pertaining to the substance described?
V06	Was the scientific methodology used in the report accepted without dispute?
V07	Was the toxicology report related to the crime scene?
V08	Was the report completed within the statutory period?
V09	Was the report produced by an official expert*?
V10	Did the report explain the sampling method utilized with the material?

*In Brazil, official experts are professionals who have special training given by the State to solve criminal cases. We consider that these professionals are, *lato sensu*, more experienced in forensic analysis than *ad hoc* experts.

2.1. Qualitative Analysis

Several variables were selected; the formulation of these questions met the requirements of what is needed to clarify the law. These variables were formulated to generate simple answers, "YES" or "NO." When the answer was positive, such an event was considered favorable to the quality of the report. Similarly, a negative response was assessed as being unfavorable in the case study, which indicated a weakness in the document. A different set of conditions was designed for preliminary and final reports. They are summarized in Table 1 and Table 2.

The analysis was performed by a single person, after training and standardization of the positive and negative requirements to assess the quality of the reports.

Table 2. Variables studied in the analysis of final reports.

V01	Was the report concluded if the substance was illegal?
V02	Was there a photograph attached of the material seized?
V03	Was there some information about the amount (mass) of the substance?
V04	Was the substance physically described?
V05	Were packaging and wrappers pertaining to the substance described?
V06	Was the final report qualitative?
V07	Did the report relate the purity of substance?
V08	Were the other constituents of the material evaluated cited?
V09	Were there images (Figures and Graphics) regarding to the results of the analysis?
V10	Did the expert who took part in the preliminary analysis also participate in the final report?
V11	Did the report evaluate materials found at the crime scene? (scales, bottles, etc.)
V12	Did the final report confirm the preliminary regarding the nature of the substance?
V13	Was the report completed within the statutory period?
V14	Was the report able to provide enough information to assist the judge to decide on the difference between user and dealer?

2.2. Quantitative Analysis

The answers obtained in the previous step produced a binary matrix: a positive response represented by "1" and a negative answer by "0". The parameters were developed to provide an empirical equation for Report Relevance **(RR).** This equation was intended to provide a quantitative indication of the amount of the information taken into account in each report.

$$RR = \frac{\sum_{i=1}^{n} Wv(i)Fc(i)Vq(i)}{\sum_{i=1}^{n} Wv(i)Fc(i)} \quad (1)$$

Where:

W_v **=Variable Weight:** the main purpose of this parameter is to give relative importance for each variable, *i.e.*, it was thought to correct distortions regarding the importance of each variable. A number was assessed to represent the relative weight of each variable. When the variable was considered to be relevant, its value was set as 1. For necessary considerations, the weight was considered as 2; for fundamental questions, the variable was assigned as 3. Table 3 and Table 4 summarize the adopted values and list explanations of these values in the case of each variable for preliminary and final reports, respectively.

Context factor (F_c): this parameter is intended to ponder the variables considering the context of the criminal action. It is specific to each report and provides a more sensitive analysis, because the situation can affect the relevance of the variables. *Fc* values were 0 for “irrelevant”, which means that the answer does not apply to the studied case. For example, we can consider a report whose goal was to determine the chemical profile of various drugs, from different police cases. In such a situation, as the date of each seizure is different, the question about the report being completed within the statutory period (V08, Table 1 and V13, Table 2) becomes “irrelevant”. The score was set as 1 for relevant; 2 for necessary and 3 when it was considered fundamental.

Vq is the variable of the question (sum of answers to the formulated variables).

Equation 1 will always provide RR values between the ranges of 0 to 1. If all variables returned "NO" responses, ***Vq*** values would result in 0, which would make RR = 0.

Table 3. Variable weight values and reasons for relevance for preliminary reports.

Variable	Weight	Reason for revelevance
V01	Fundamental (3)	The motive for the arrest of the accused.
V02	Necessary (2)	Aids in identifying the substance.
V03	Fundamental (3)	The law requires an indication of the amount of the seized substance.
V04	Relevant (1)	Despite being important, the absence of this information does not prejudice the rest of the report.
V05	Relevant (1)	Can influence the characterization of the offense.
V06	Necessary (2)	Despite the importance of specificity in the preliminary test, sensitivity is most valued; the nature of the substance will be confirmed by the final report, with appropriate methodologies.
V07	Relevant (1)	The absence of this information does not affect the progress of the case.
V08	Necessary (2)	Depending on the situation, the late delivery of the report may cause the nullity of the criminal proceedings.
V09	Necessary (2)	*In Brazil, official experts are professionals who have special training giving by the State to solve criminal cases. We consider that these professionals are, *lato sensu*, more experienced in forensic analysis than *ad hoc* experts.
V10	Relevant (1)	In practice, the most common seizures are small amounts.

Table 4. Weight values and reasons for relevance for final reports.

Variable	Weight	Reason for revelevance
V01	Fundamental (3)	This information provides the basis for determining the existence of the offense.
V02	Necessary (2)	Assists in the interpretation of the report.
V03	Relevant (1)	This is a procedure that does not directly influence the judge's decision.
V04	Relevant (1)	Despite its relative importance this has been classified as only relevant as it does not prejudice the rest of the report.
V05	Fundamental (3)	In the final report, the methodology must be robust in order to provide legal certainty.
V06	Fundamental (3)	Even though nature of the substance is addressed in the preliminary report, it is of crucial importance to be confirmed in the final exam.
V07	Relevant (1)	In practice, this question is rarely raised.
V08	Relevant (1)	This does not change the decision based on the exam.
V09	Necessary (2)	Gives greater strength to the test results.
V10	Relevant (1)	There is no need for the same expert to conduct both tests.
V11	Relevant (1)	In some cases, this may be important.
V12	Necessary (2)	In case of discrepancy, there may have been a methodological error in the preliminary report.
V13	Necessary (2)	This can invalidate the criminal process.
V14	Fundamental (3)	Can assist the judge in determining the sentence.

Similarly, if all variable responses were "YES", the ***Vq*** values would be equal to 1.

The relevance of the report can be interpreted as a probability related to the amount of information involved in the document. If a report presents a RR = 1, we can say that it probably contains 100% of the minimum information necessary for a good interpretation of the criminal activity. On the other hand, if RR = 0, it can be inferred that this report does not lend itself as evidence in a court of law. Intermediate values, in turn, can be interpreted in accordance with the amount of information associated with a particular context.

2.3. Data Processing

The binary matrices were built in Microsoft Office Excel®. With the use of programming cells, Equation 1 was inserted in each cell of the matrix, and the relevance of the preliminary and final reports was obtained (Figure 1).

Preliminary report			
Variables	Wv	Fc	Answers
V01	3		
V02	2		
V03	3		
V04	1		
V05	1		
V06	2		
V07	1		
V08	2		
V09	2		
V10	1		

Final report			
Variables	Wv	Fc	Answers
V01	3		
V02	2		
V03	1		
V04	1		
V05	3		
V06	3		
V07	1		
V08	1		
V09	2		
V10	1		
V11	1		
V12	2		
V13	2		
V14	3		

Preliminary	
Total score	RR

Final	
Total score	RR

Figure 1. Example of cells used to calculate RR for each pair of reports.

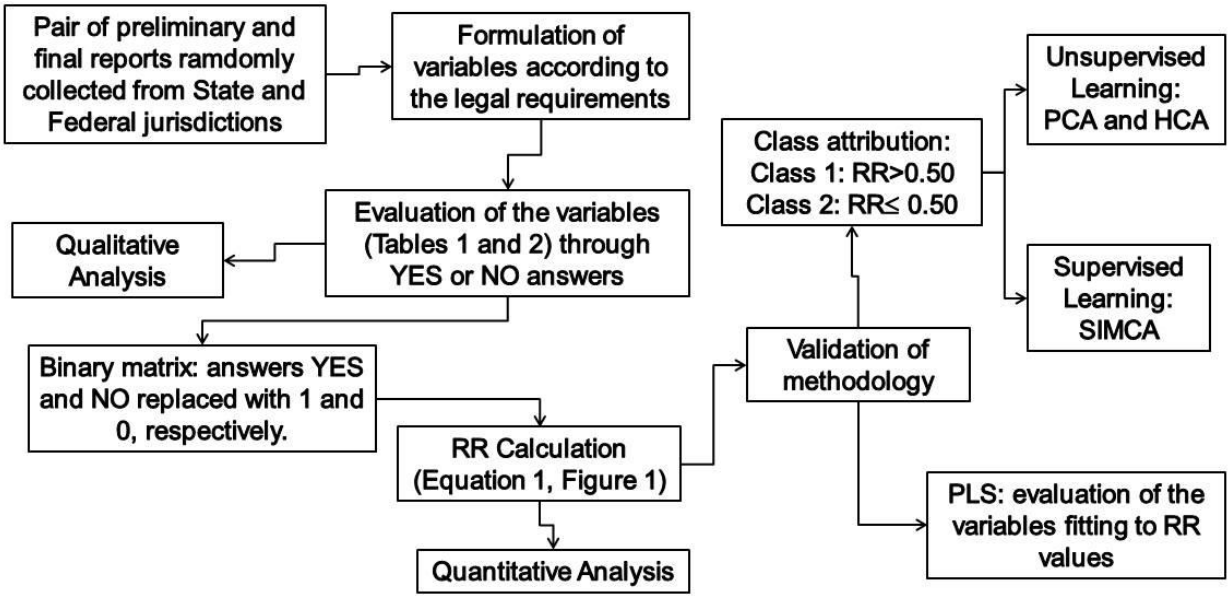


Figure 2. Scheme for reports evaluation.

To verify the acceptability of the proposed analytical methodology, it was necessary to validate it. The official documents were assigned into two distinct classes:

- Class 1 - reports with RR greater than 0.50 (RR> 0.50), which means that the document in question contained more than 50% of the necessary information.
- Class 2 - reports with RR less than or equal to 0.50 (RR ≤ 0.50), which means that the document in question contained less than half of the necessary information.

Multivariate Analysis was used to validate the current methodology, using the Pirouette® package. [18–21] Pattern recognition techniques were used to verify how the samples were clustered. Unsupervised learning methods PCA and HCA were used to verify natural similarities, while the supervised learning method, SIMCA, was used to evaluate the efficacy of the previous classification.

The PLS technique was used to measure the tuning of the variables chosen through empirical equation. The idea is to understand how the variables fit to RR values. Figure 2 presents a scheme used in the process.

3. RESULTS

3.1. Qualitative Analysis

3.1.1. Preliminary reports

Table 5 presents results for preliminary reports. Red values indicate variables with results corresponding to reports with negative responses. V01, V03, V04, V05 and V09 had positive responses for almost all the reports. V08 showed "YES" in most of the cases. V02 and V10 had low frequency of positive answers. V07 returned a negative response in almost all cases.

Only V06 returned 100% of negative responses, revealing that the tests used in the provisional analysis of narcotics are fragile, which may generate discrepancies. With respect to information with a higher degree of importance, such as the nature of the substance (V01) and quantity (V03), percentages were generally positive.

3.1.2. Final reports

Table 5. Responses for preliminary reports.

Variables	State Reports		Federal Reports		Total	
	YES	NO	YES	NO	YES	NO
V01	97,96%	2,04%	96,97%	3,03%	97,83%	2,17%
V02	17,54%	82,46%	78,46%	21,54%	25,40%	74,60%
V03	91,12%	8,88%	95,38%	4,62%	91,67%	8,33%
V04	96,36%	3,64%	98,46%	1,54%	96,63%	3,37%
V05	93,62%	6,38%	95,38%	4,62%	93,85%	6,15%
V06	0,00%	100,00%	0,00%	100,00%	0,00%	100,00%
V07	0,46%	99,54%	20,00%	80,00%	2,98%	97,02%
V08	66,97%	33,03%	92,31%	7,69%	70,24%	29,76%
V09	97,72%	2,28%	100,00%	0,00%	98,02%	1,98%
V10	37,81%	62,19%	95,38%	4,62%	44,84%	55,16%

Table 6 shows the results for the final reports. Most variables returned positive responses. There were inferior results for variables V07, V08, V09, V10 and V11. For the most relevant variables, V01, V05, V06 and V14 (Table 4) good rates of positivity were obtained and V06 had 100% of "YES" answers. The best results were for the reports from Federal Police. In general, the reports had a good level of standardization.

3.2. Quantitative analysis

3.2.1. Report Relevance

Table 7 presents results for RR, Average, Standard Deviations (SD), Maximum and Minimum values found for both preliminary and final reports from the different jurisdictions.

For State Jurisdictions, preliminary and final reports returned, respectively, an average value of 0.70 ± 0.10 and 0.75 ± 0.10 respectively. This is a good representative importance. Although the standard deviation range is around 9%, we observed that maximum and minimum values varied considerably, which indicates great discrepancies in the documents. For final reports, these differences were less critical.

For Federal Jurisdictions, RR values for both preliminary and final reports were significantly higher – average values 0.83 and 0.89, respectively; the standard deviations (SD) were also lower – around 7%, as well as the difference in minimum and maximum values.

These results are as expected, since the Federal Police is the point of reference for State Police forces; it receives greater investment in technology and training, and has a lower workload. These factors enable Federal Police experts to perform their task more adequately.

About the overall results, we can say that the scores were good because the average values were higher than 0.7 for both preliminary and final documents. However, the standard deviations (SD) were around 10%, higher than those found for each jurisdiction. State reports were responsible for the low scores found for minimum values for both preliminary and final documents.

The frequencies for RR values relating to preliminary reports showed that for State jurisdictions, more than 88% have scores over 0.63. For Federal documents, almost 90% have high scores (over 0.74). The total amount of the documents showed a good score, since around 80% of them have RR values over 0.64.

For final reports, values were better, since around 68% of State documents have values over 0.77; for Federal ones, the results are even better, since around 86% of them have scores over 0.85. Evaluating the total RR values, we find that almost 70% of them have values over 0.79.

Table 6. Responses for final reports.

Variables	State Reports		Federal Reports		Total	
	YES	NO	YES	NO	YES	NO
V01	97,28%	2,72%	98,48%	1,52%	97,44%	2,56%
V02	53,99%	46,01%	73,85%	26,15%	56,55%	43,45%
V03	95,67%	4,33%	100,00%	0,00%	96,23%	3,77%
V04	100,00%	0,00%	100,00%	0,00%	100,00%	0,00%
V05	78,13%	21,87%	90,77%	9,23%	79,76%	20,24%
V06	100,00%	0,00%	100,00%	0,00%	100,00%	0,00%
V07	0,00%	100,00%	0,00%	100,00%	0,00%	100,00%
V08	0,46%	99,54%	35,38%	64,62%	4,96%	95,04%
V09	0,23%	99,77%	86,15%	13,85%	11,31%	88,69%
V10	16,86%	83,14%	10,77%	89,23%	16,07%	83,93%
V11	1,37%	98,63%	10,77%	89,23%	2,58%	97,42%
V12	97,51%	2,49%	95,45%	4,55%	96,48%	3,52%
V13	55,58%	44,42%	87,69%	12,31%	59,72%	40,28%
V14	91,57%	8,43%	100,00%	0,00%	92,66%	7,34%

Table 7. Reports Relevance according to Equation 1.

	RR values	Preliminary´s RR	Final´s RR
State Reports	Media	0.70	0.75
	SD	0.10	0.10
	Maximum	0.87	0.86
	Minimum	0.26	0.41
Federal Reports	Media	0.83	0.89
	SD	0.07	0.08
	Maximum	0.90	0.95
	Minimum	0.63	0.71
Total	Media	0.72	0.77
	SD	0.11	0.11
	Maximum	0.90	0.95
	Minimum	0.26	0.42

3.2.2. Validation of Methodology

In this section, the main goal is to apply multivariate analysis to validate Equation 1 according to the values chosen for its parameters W_v and F_c (Tables 3 and 4) and the variables of the questions formulated to evaluate the reports. In these cases, all samples were analyzed together for preliminary and final documents, with no separation between State and Federal Reports.

Most reports were classified as Class 1. For preliminary According to unsupervised learning techniques, we observed that V03 and V08 were the variables that influenced Class 2 reports. In fact, these samples presented exclusively negative answers for these variables. On the other hand, V02, V07 and V10 are separate from the other variables. Table 5 shows that these variables returned predominantly negative answers. Table 8 shows the samples for descendants from the lower group in Figure 4(a). All of them have scores no higher than 6.5 and the respective results for variables. A combination of negative results is observed for at least three of variables V02, V03, V07, V08 and V10 for most of samples. Reports 144, 160, 161 and 441 have higher scores and only two negative answers, which is why they are located at the top of this group.

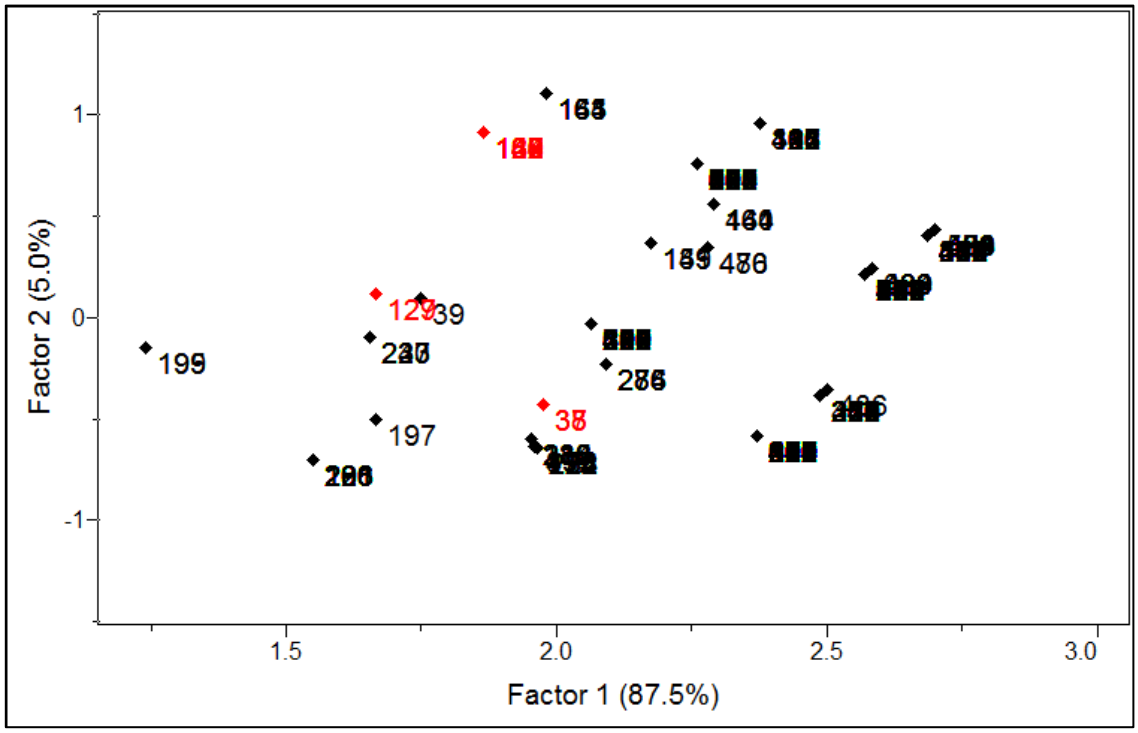
documents, 24 samples (4.8%) were assigned as Class 2. For final reports, this number was lower, since only 11 reports (2.2%) had Class 2 scores.

Unsupervised learning techniques PCA and HCA were used to evaluate clustering; SIMCA was used to evaluate clustering from the supervised point of view. Finally, PLS was performed on data to fit the variables to RR values. PLS modeling was not performed to obtain a preview modeling, since each report had to be evaluated; it was used as a tool to verify the influence of variables on the composition of RR, considered to be variable- dependent.

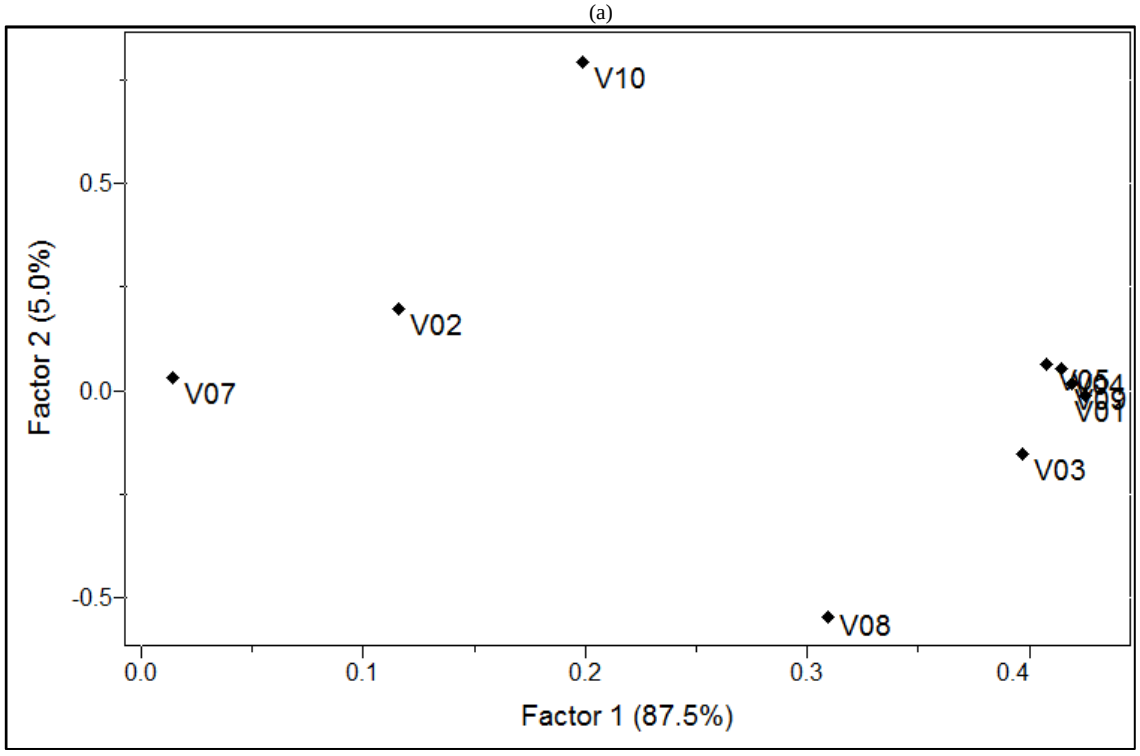
3.2.2.1. Preliminary reports

V06 was removed from original Matrix, once 100% of the answers were NO and consequently this variable did not contribute to determining RR values.

Unsupervised Learning: PCA and HCA results



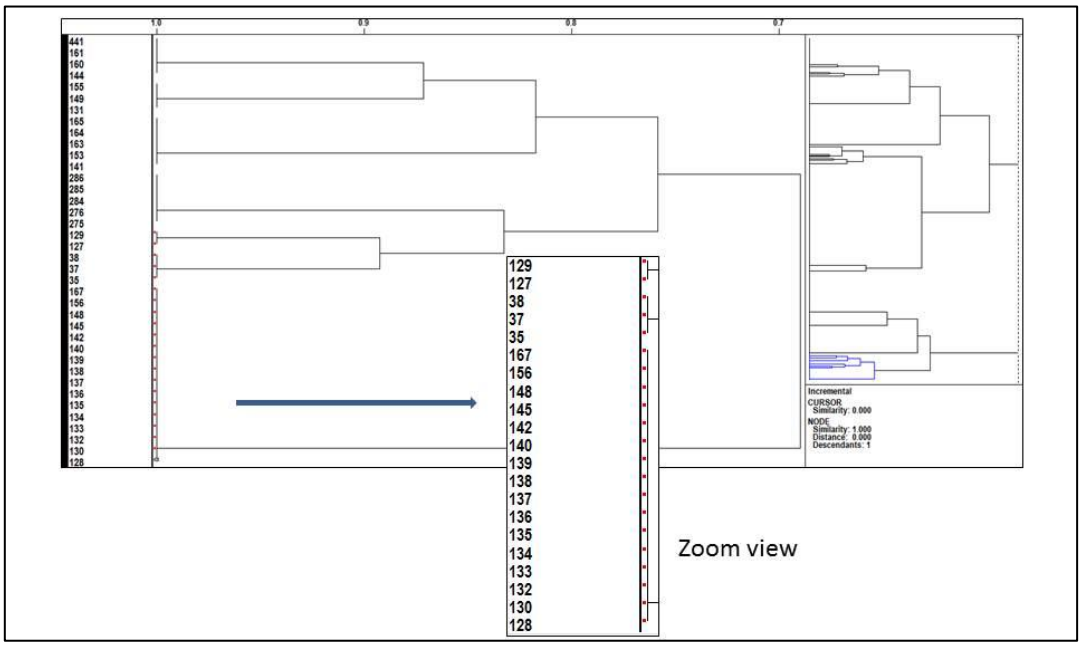
(a)



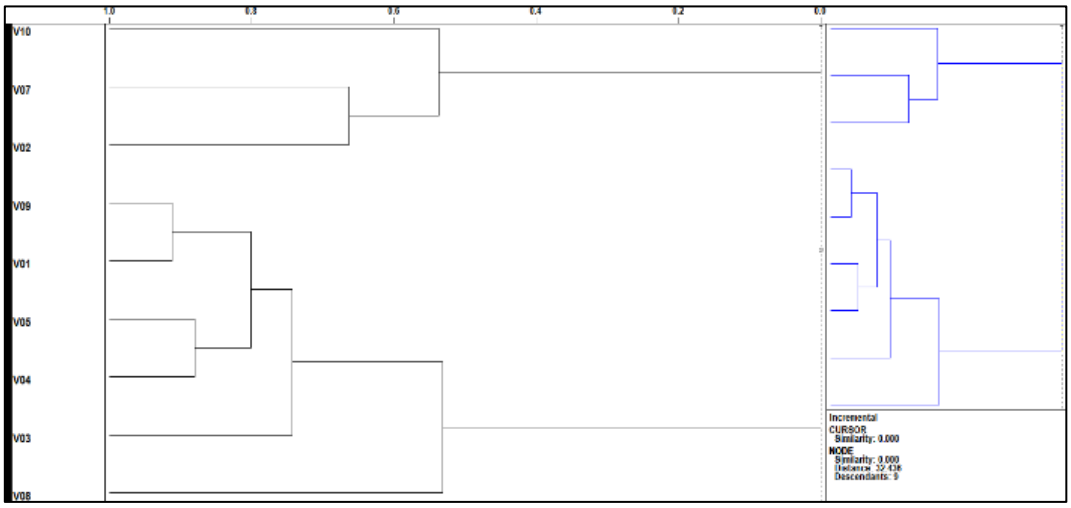
(b)

Figure 3. Pairs of Scores (a) and Loadings (b) for preliminary reports

.



(a)



(b)

Figure 4. Dendrograms for samples (a) and variables (b) for preliminary reports.

PCA was performed over the original matrix. Figure 3 shows the pair scores v. loadings: samples for Class 1 are black and for Class 2 are red. Two Factors or PCs were able to account for around 92.5% of the overall information. We observe that there was no clear separation between the samples. In this figure we observe that there are overlapping points. It means that there are reports with similar answers for most of the evaluated variables. The individual values for each report are in the supplementary material.

The HCA was performed on the data using Euclidian distance and incremental linkage method. Figure 4 shows the dendrograms for the preliminary documents.

SIMCA Results

Three PCs were found in the analysis. The interclass distances were 1.94; these values are higher than the used cutoff, *i.e.*, 0.5. Table 9 shows the interclass residuals, which presented greater values for the other Class than those found for the Class itself.

Table 8. Data for descendants of the lower group in HCA for preliminary reports.

	V02	V03	V07	V08	V10	RR
35	0	0	0	1	0	0.5000
37	0	0	0	1	0	0.5000
38	0	0	0	1	0	0.5000
127	0	0	0	0	0	0.3947
129	0	0	0	0	0	0.3947
130	0	0	0	0	1	0.4211
131	0	0	0	1	1	0.5263
132	0	0	0	0	1	0.4211
133	0	0	0	0	1	0.4211
134	0	0	0	0	1	0.4211
135	0	0	0	0	1	0.4211
136	0	0	0	0	1	0.4211
137	0	0	0	0	1	0.4211
138	0	0	0	0	1	0.4211
139	0	0	0	0	1	0.4211
140	0	0	0	0	1	0.4211
141	1	0	0	0	1	0.5263
142	0	0	0	0	1	0.4211
144	**1**	**0**	**0**	**1**	**1**	**0.6316**
145	0	0	0	0	1	0.4211
148	0	0	0	0	1	0.4211
149	0	0	0	1	1	0.5263
153	1	0	0	0	1	0.5263
155	0	0	0	1	1	0.5263
160	**1**	**0**	**0**	**1**	**1**	**0.6316**
161	**1**	**0**	**0**	**1**	**1**	**0.6316**
163	1	0	0	0	1	0.5263
164	1	0	0	0	1	0.5263
165	1	0	0	0	1	0.5263
275	1	0	0	1	0	0.6053
276	1	0	0	1	0	0.6053
284	1	0	0	1	0	0.6053
285	1	0	0	1	0	0.6053
286	1	0	0	1	0	0.6053
441	**1**	**0**	**0**	**1**	**1**	**0.6316**

Figure 5 shows the separations of Classes. For preliminary reports, three of them (131, 149 and 155) initially assigned as Class 1 were classified as Class 2, despite having a RR value higher than 0.50. This value represents around 0.61% of the total samples. They presented negative answers for both V02 and V03. For samples initially assigned as Class 1, 24 were unclassified (39, 188, 193, 195-197, 199-201, 203, 206, 208, 221, 223, 231, 233, 236, 237, 240, 394, 486, 488, 489, 499). According to the literature, a model which classifies 95% of samples correctly is highly trustworthy.[21]

Table 9. Interclass residuals for preliminary reports.

	Class 1	Class 2
Class 1	0.19	0.44
Class 2	0.34	0.10

Table 12 shows the values for Discriminating and Modeling powers. The Discrimination Power is a measure used to determine the influence of each of the variables in the separation of classes. The higher the value, the greater the ability of a variable to discriminate between samples in different classes. The modeling power, in turn, is related to the importance of each variable in describing the information from the training set for each class.[21,22] Variables V02 and V03 have a higher discriminating

power, while V02, V08 and V10 show the higher modeling power.

PLS Results

To verify the fitting of the variables to RR values, a PLS regression was performed. Three PCs were found, accumulating 93.6% of the information. The following parameters were evaluated:

a) Correlation coefficient for cross validation Q^2 = 0.98021
b) Correlation coefficient for calibration R^2= 0.98211
c) Values for Root Mean Square error for cross validation (RMSEV) 0.016841
d) and Root Mean Square error for calibration (RMSEC) 0.016233

The results obey the conditions for a strength modeling: $Q^2<R^2$ and RMSEV< RMSEC.[23,24] Cross validation was performed removing from one to 51 samples (Leave One Out (LOO) until Leave 51 Out (LNO, N=51)). The idea was to observe the behavior of the modeling by removing more than 10% of the samples. We observed the same results for all attempts, which assured us of the robustness of the modeling. Figure 6 shows the PLS calibration curve. Numerical values can be conferred in Table 3.

Table 11 shows the regression vector: the most important variables for describing RR values were V01 and V03, related to the nature and amount of substance, respectively. These questions are essential in describing the criminal offense.

Table 11. Coefficients for each Variable contributing to RR in preliminary reports.

	Coefficient
V01	**0.144048**
V02	0.124429
V03	**0.226456**
V04	0.064681
V05	0.041594
V07	0.017320
V08	0.117869
V09	0.139833
V10	0.030984

3.2.2.2. Final reports

In this case, V07 was removed from the original Matrix (Table 2). Table 6 shows that 100% of the answers were negative, which indicates that V07 made no contribution to RR values.

Unsupervised Learning: PCA and HCA results

Three PCs (Factors) accounted for approximately 95% of the information as a whole. Figure 7 shows the bi-dimensional distribution for scores and loadings. In this case, we also have overlapping points meaning that many reports have similar answers for the variables. The individual values for each report are in the supplementary material.

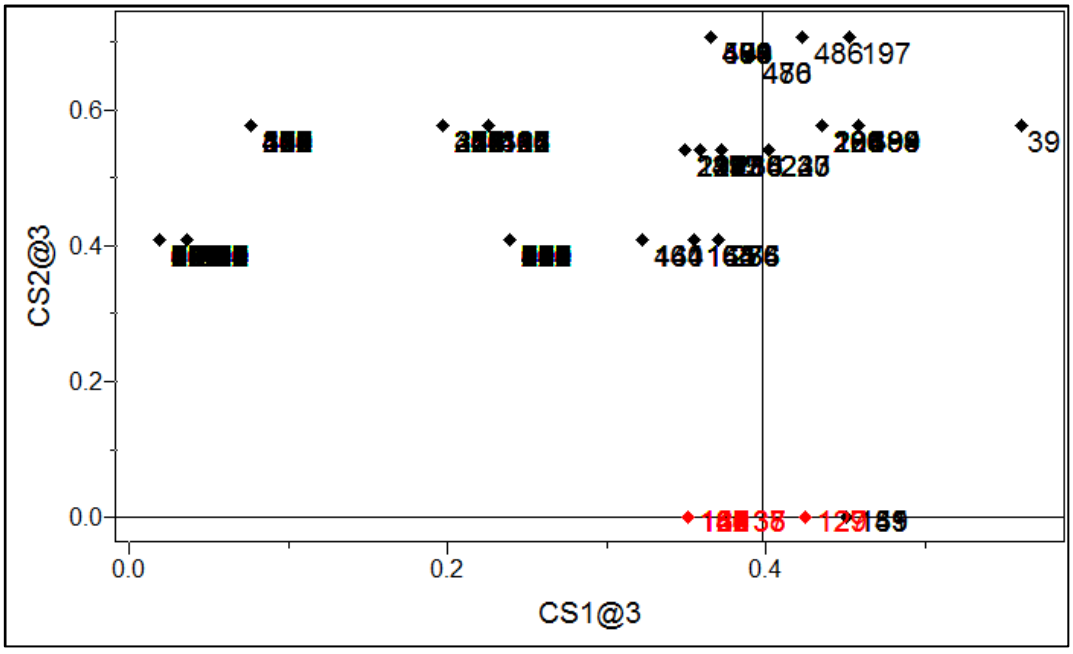


Figure 5. SIMCA Modeling with 3PCs.

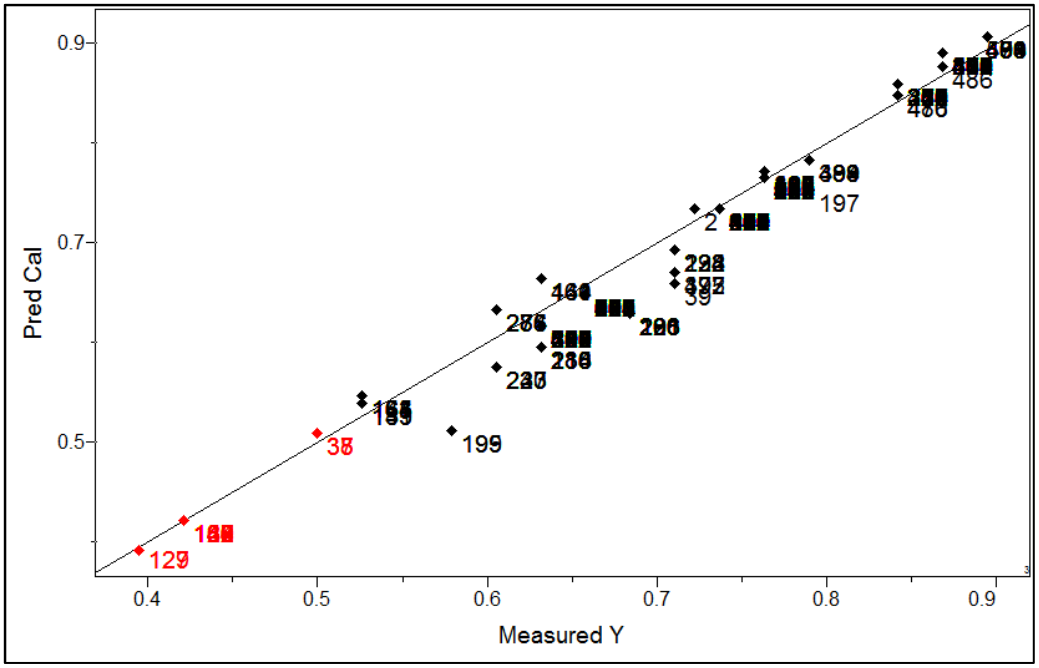
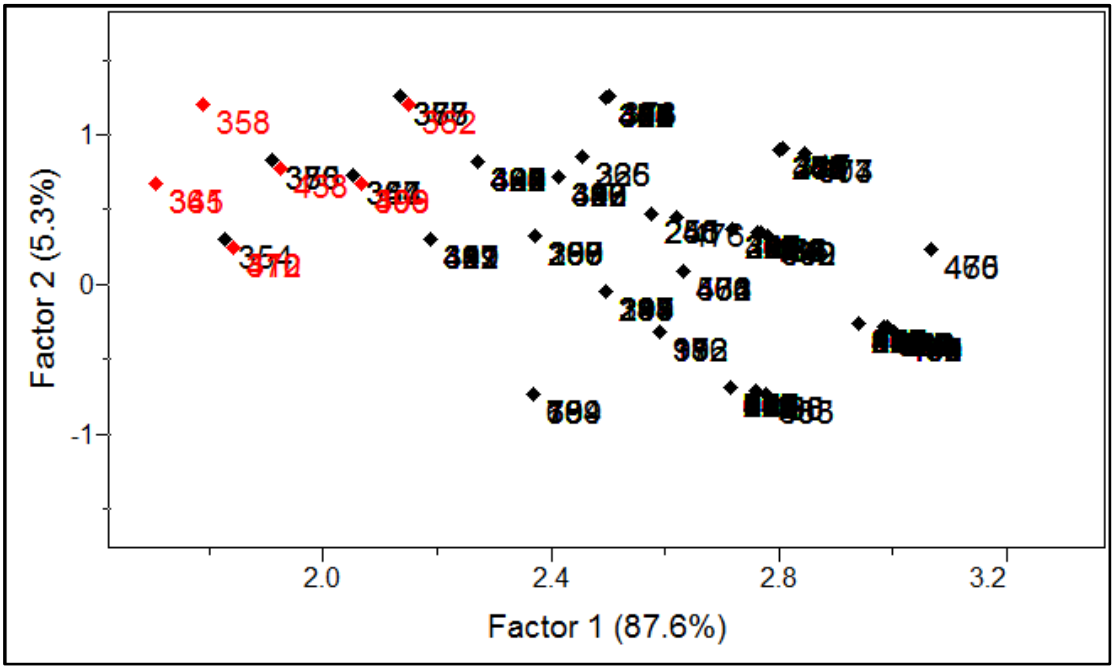


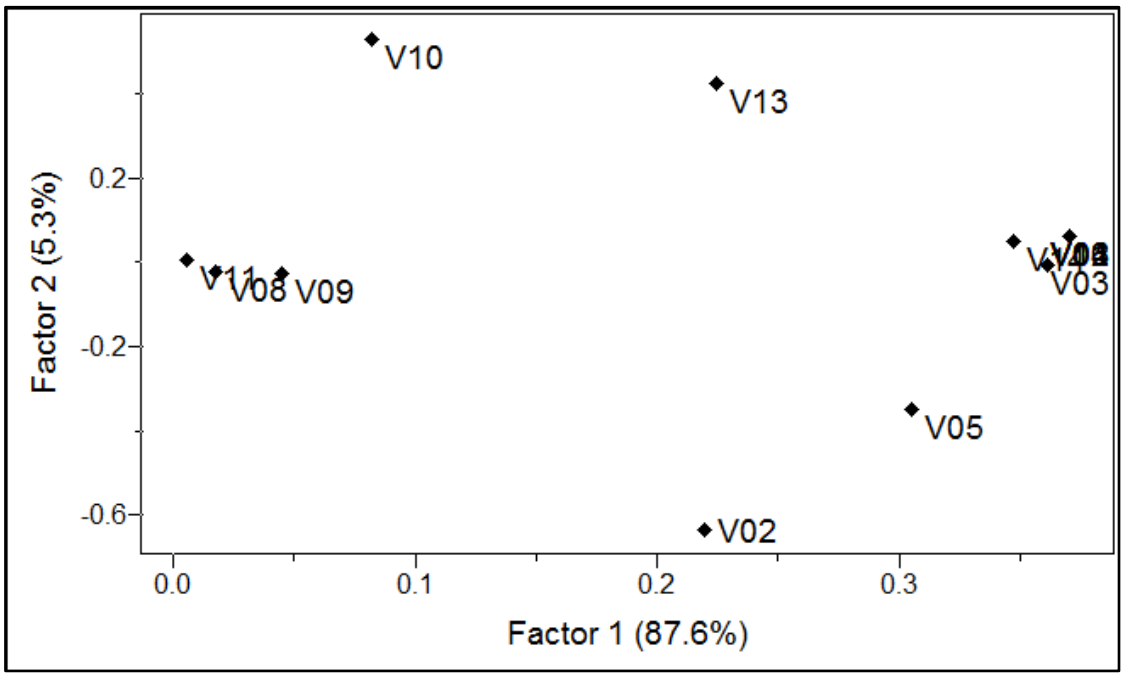
Figure 6. PLS Calibration Plot for preliminary reports.



(a)

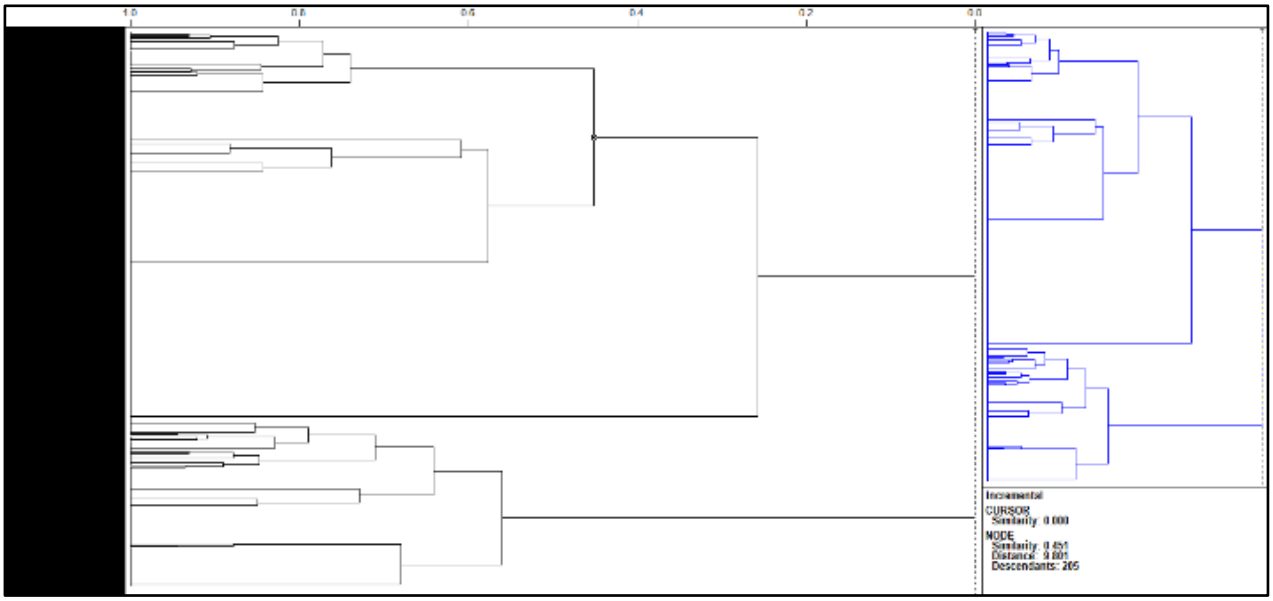
Figure 8 presents the HCA results for samples (a) and variables (b).

Both PCA and HCA showed some overlap for samples from Class 1 and Class 2. The HCA lower group in Figure 8(a) has 148 descendants including the samples with RR ≤ 0.50. This overlap is largely due to the fact that these samples have similar results for the lower group (V08-V11). There are at least three negative answerers for these variables. In addition, they also have at least one negative value for variables V02, V05 and V13. The same behavior can be observed for PCA Loadings (Figure 7b).



(b)

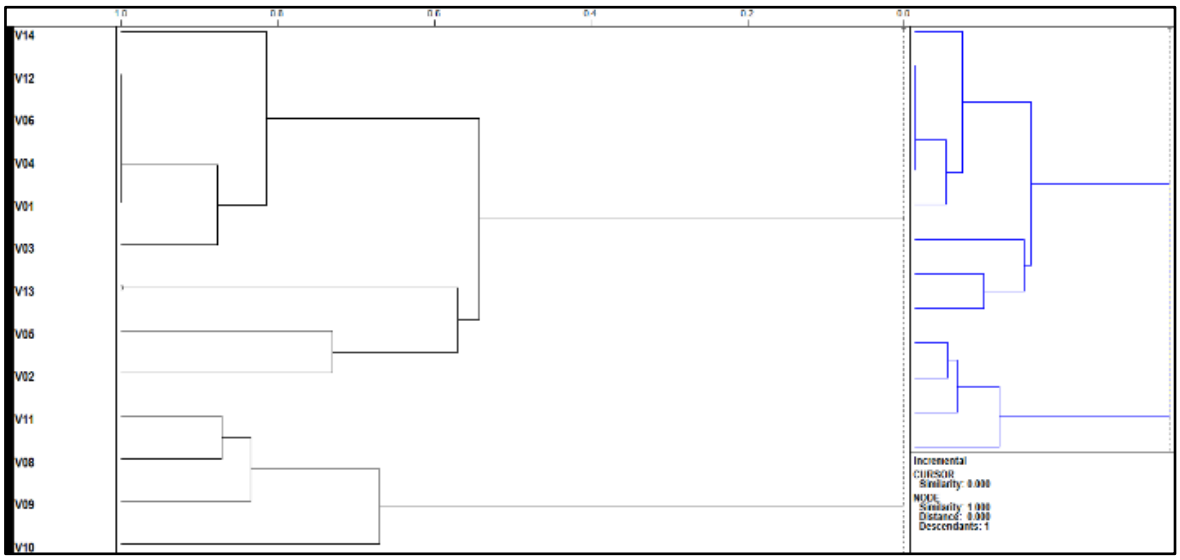
Figure 7. Pair of Scores (a) and Loadings (b) for final reports.



Lower Group- 148 descendants

197, 219, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 264, 266, 270, 271, 272, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 287, 288, 293, 294, 295, 297, 302, 303, 304, 305, 306, 314, 315, 317, 331, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 372, 373, 374, 375, 376, 378, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 395, 396, 397, 398, 399, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 460, 463, 466, 472, 486, 501

(a)



(b)

Figure 8. Dendograms for samples (a) and variables (b) for final reports.

SIMCA Results

Three PCs were chosen for modeling and the interclass distance was around 1.99, higher than the 0.5 cutoff. Table 12 shows the interclass residuals, which have greater values for the other class than for the class itself. Class separations are shown in Figure 9. No sample was misclassified and 14 of them, firstly assigned to Class 1, were not identified in any previous classification (448, 449, 452, 465, 470, 476, 485, 488, 489, 492, 496, 498, 500, 503). This accounts for 2.87% of the samples, confirming the good modeling for SIMCA, since more than 95% of the reports were correctly classified.[21] Values for Discriminating and Modeling powers are shown in Table 13.

Table 12. Interclass residuals for final reports.

	Class 1	Classe 2
Classe 1	0,19	0,49
Classe 2	0,35	0,09

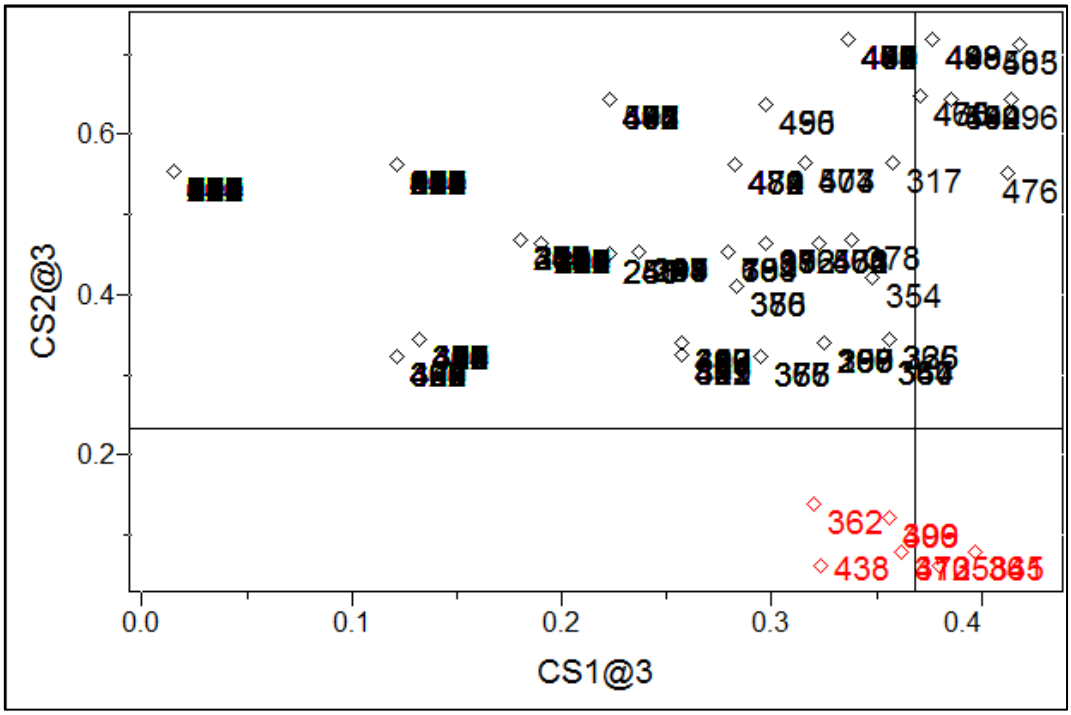


Figure 9. SIMCA Modeling with 3PCs for final reports.

Table 13. Variable Discrimination and Modeling Powers for final reports.

	Discriminating Power	Total Modeling Power
V01	2.284011	1.000000
V02	**13.483626**	0.660987
V03	2.670452	0.584790
V04	2.284000	1.000000
V05	**13.928339**	0.499850
V06	2.283999	1.000000
V08	1.222491	0.252063
V09	1.555229	0.311006
V10	0.065266	0.749601
V11	1.036783	0.200475
V12	2.283999	1.000000
V13	3.423817	0.700157
V14	**23.470266**	0.293845

Three PCs were selected for the model, accounting for around 94% of the information as a whole. The parameters were found to obey the conditions for a strength modeling: $Q^2<R^2$ and RMSEV>RMSEC:[23,24]

a) Correlation coefficient for cross validation Q^2 = 0.97759;
b) Correlation coefficient for calibration R^2= 0.97872;
c) Root Mean Square error for Validation (RMSEV): 0.015571;
d) Root Mean Square error for Calibration (RMSEC): 0.015228.

Cross validation was performed removing from one to 51 samples (Leave One Out (LOO) until Leave 51 Out (LNO, N=51)). The same results were observed in all cases. The PLS calibration curve is shown in Figure 10, and correspondent numerical values are given in Table 4.

PLS Results

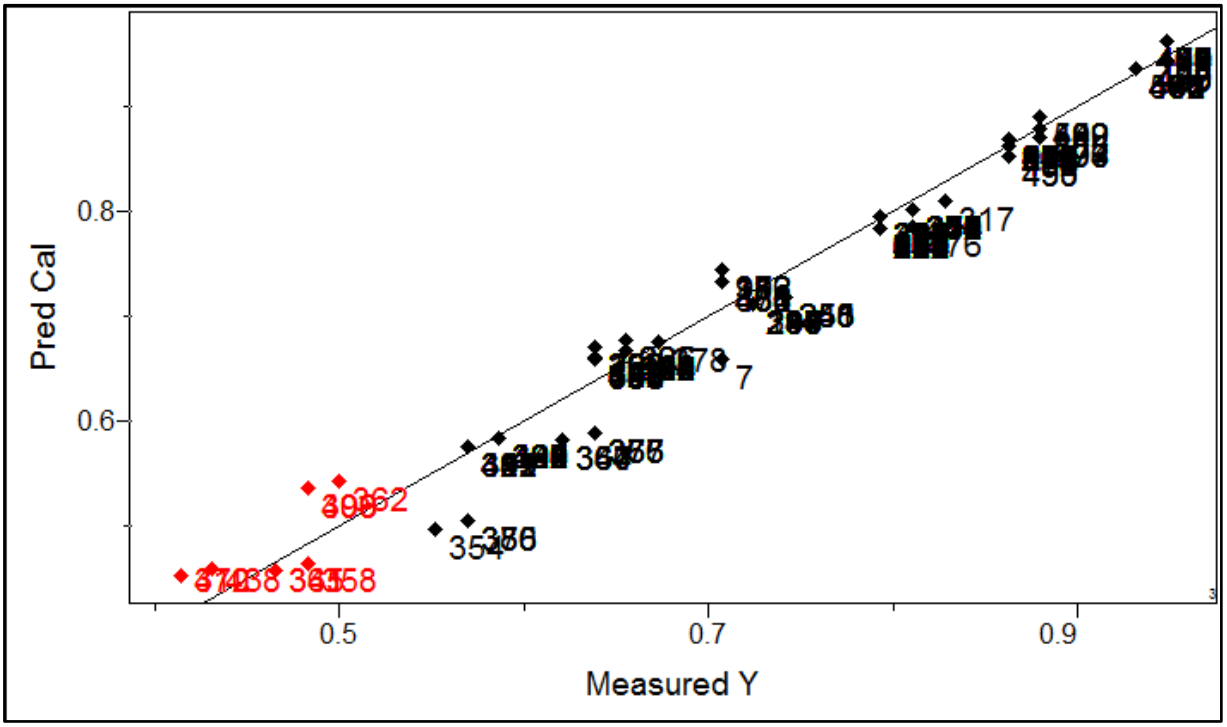


Figure 10. PLS Calibration Plot for final reports.

Table 14 shows the regression vector: the most important variables for describing RR values were V01 and V03, related to the nature and amount of substance, respectively. These questions are essential for describing the criminal offense.

Table 14. Coefficients for each Variable contributing to RR values for final reports.

Variables	Coefficients
V01	0.093262
V02	0.073046
V03	0.078877
V04	0.093262
V05	**0.134368**
V06	0.093262
V08	0.026795
V09	0.067995
V10	0.007091
V11	0.007974
V12	0.093262
V13	0.084179
V14	**0.124290**

V05 and V14 were the most important variables contributing to RR values. V05 is related to the evaluation of packing and wrappers, confirming the importance of this procedure. V14 is related to the ability of the report to provide the necessary information to assist the judge in deciding if the accused is a user or a dealer. This is extremely important, since the correct application of the penalties is a question of human rights.

4. DISCUSSION

Our qualitative analysis showed that, in general, forensic reports on illicit drugs seized by the police are reliable, since most of their variables return with YES answers for both preliminary and final reports.

Quantitative analysis confirmed this finding. The RR values were on average 0.74 ± 0.08 for the state reports. Federal reports obtained average RR values of 0.87 ± 0.05.

Validation of the methodology showed that unsupervised learning was able to separate similar reports according to variable results, but did not provide a good separation of the classes. On the other hand, SIMCA classification was successful in discriminating between the classes according to the previous training set. Finally, for both preliminary and final reports, we found high values for Q^2 and R^2, which confirms that the variables chosen were well fitted to RR values. The cross validation results from LOO to LNO, N=51 showed that the PLS modeling is robust in both cases.

5. CONCLUSIONS

In the present study, we applied a methodology in order to assess expert reports on illicit drugs seized by the police qualitatively and quantitatively. The empirical equation used was validated by multivariate data analysis, which was effective for this purpose. The variables used in this evaluation were questions based on what is necessary for law enforcement. The quality of the reports was considered good according to RR values. With strengths and weaknesses, each group of reports studied had its own model utilized by the expert responsible. This relative uniformity of the reports, however, revealed some weaknesses, which offer scope for improvement. Some final reports utilized analytical methodologies in non-compliance with international recommendations. This may be one of the causes of non-agreement between the preliminary and definitive determination of the drug, with false-positive, false-negative or indeterminate results. Another important observation was that most of the reports had no accompanying images, an important element in aiding judges to understand the reports. The insertion of photographs is easy to perform and enhances the quality of the expert work. Finally, our methodology has the potential for use in the analysis of forensic reports on and other types of case, which opens the prospect of further studies.

6. ACKNOWLEDGMENTS

We thank the Federal and Forensics Centers and Expertise Nucleus for collaboration with this work and provision of the analyzed reports. The authors thank the Brazilian Agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (ATB and JAV: INCT Forense-CNPq, Grant 465450/2014-8,), Fundação de Amparo à Pesquisa do Estado de São Paulo (VBPL: FAPESP, Grant:2016/19766-1) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

REFERENCES

[1] I. Dror. The ambition to be scientific: Human expert performance and objectivity., Sci. Justice. **53**, 81–2, 2003.

[2] R. Robinson, M.W. Acklin, Fitness in paradise: Quality of forensic reports submitted to the Hawaii judiciary, Int. J. Law Psychiatry. **33**, 131–137, 2010. doi:10.1016/j.ijlp.2010.03.001.

[3] K.D. Fuger, M.W. Acklin, A.H. Nguyen, L. a. Ignacio, W.N. Gowensmith, Quality of criminal responsibility reports submitted to the Hawaii judiciary, Int. J. Law Psychiatry. **37**, 272–280, 2014. doi:10.1016/j.ijlp.2013.11.020.

[4] N. Duits, S. van der Hoorn, M. Wiznitzer, R.M. Wettstein, E. de Beurs, Quality improvement of forensic mental health evaluations and reports of youth in the Netherlands, Int. J. Law Psychiatry. **35**, 440–444, 2012. doi:10.1016/j.ijlp.2012.09.018.

[5] A. Loughnan, T. Ward, Emergent authority and expert knowledge: Psychiatry and criminal responsibility in the UK, Int. J. Law Psychiatry. **37**, 25–36, 2014. doi:10.1016/j.ijlp.2013.10.002.

[6] J. Goodman-Delahunty, M.K. Dhami, A Forensic Examination of Court Reports, Aust. Psychol. **48**, 32–40, 2013. doi:10.1111/j.1742-9544.2012.00082.x.

[7] A. Kloosterman, M. Sjerps, A. Quak, Error rates in forensic DNA analysis: Definition, numbers, impact and communication, Forensic Sci. Int. Genet. **12**, 77–85, 2014. doi:10.1016/j.fsigen.2014.04.014.

[8] J.A. Bright, J.M. Curran, J.S. Buckleton, The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation, Forensic Sci. Int. Genet. **12**, 208–214, 2014. doi:10.1016/j.fsigen.2014.06.009.

[9] KLINKNER, Melanie. Forensic science expertise for international criminal proceedings: an old problem, a new context and a pragmatic resolution. The International Journal of Evidence & Proof, **13(2)**, 102-129, 2009. doi:10.1350/ijep.2008.12.1.284.

[10] C. Maria, S.D.E. Santana, A PERÍCIA CONTÁBIL E SUA CONTRIBUIÇÃO NA SENTENÇA JUDICIAL : um estudo exploratório, Universidade de São Paulo, 1999. http://www.teses.usp.br/teses/disponiveis/12/12136/tde-06102004-161123/pt-br.php.

[11] S. Caldeira, A Influência do laudo pericial contábil na decisão dos juizes em processos nas varas cíveis., (n.d.). http://www.e.fernando.cse.prof.ufsc.br/Tema 6-A INFLUENCIA DO LAUDO PERICIAL.pdf (accessed August 5, 2014).

[12] M.F. Zochio, Qualidade dos laudos periciais emitidos na comarca de São Paulo, Universidade de São Paulo, 2010.

[13] S.C. Martinez, A Efetividade do Laudo Pericial – Um Estudo sobre a Influência do Laudo Pericial de Informática na Decisão Judicial, Fundação Getúlio Vargas, 2013.

http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/11357/VersãofinalDissertacaoSimone.pdf?sequence=1&isAllowed=y.

[14] Lei nº 11.343, DE 23 DE AGOSTO DE 2006, Presidência Da República Casa Civil, Subchefia Para Assuntos Jurídicos. (2006). http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/l11343.htm (accessed May 19, 2015).

[15] H.C. Lee, E.M. Pagliaro, Criminalistics, in: S.M. Sanbar, M.H. Firestone, S. Fiscina, T.L. LeBlang, J.W. Snyder, C.H. Wecht, et al. (Eds.), Leg. Med., 7th ed., Mosby, Inc, Philadelphia, 635–644, 2007. http://www.ablminc.org/Model_Curriculum_LMME_2010/BOOK_Legal Medicine-7th_2007/Ch68-Criminalistics.pdf.

[16] M.F. Neves, F. Rosado, Sobre a Crescente Importância da Estatística na Ciência Forense, CEAUL-Centro Estatística E Publicações Da Univ. Lisboa. (2008). http://www.ceaul.fc.ul.pt/notas.html?ano=2009 (accessed April 8, 2014).

[17] M.F. Neves, A Estatística no Tribunal, CEAUL-Centro Estatística E Publicações Da Univ. Lisboa. 1–14, 2009. http://www.ceaul.fc.ul.pt/getfile.asp?where=notas&id=260 (accessed April 8, 2014).

[18] L.S. Ramos, K.R. Beebe, W.P. Carey, E.S. M, B.C. Erickson, B.E. Wilson, et al., L. Scott Ramos, Kenneth R. Beebe, W. Patrick Carey, Eugenio Sfinchez M., Brice C. Erickson, Bruce E. Wilson, Lawrence E. Wangen,' and Bruce R. Kowalski* Laboratory for Chemometrics, Department, Education. (1986) 31–49.

[19] R.G. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, 1st ed., John Wiley & Sons, Ltd., West Sussex, England, 2003.

[20] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, Fifth, Prentice Hall, Harlow, Essex. UK., 2005.

[21] W. Infometrix Inc.: Woodinville, Pirouette, Multivariate Data Analysis for IBM PC Systems, (2001).

[22] M.M.C. Ferreira, Quimiometria: Conceitos, Métodos e Aplicações, 1st ed., Editora da Unicamp, Campinas, SP, n.d.

[23] Pirouette 3.01, Infometrix Inc. Woodinville, WA, (2001).

[24] A.T. Bruni, J.A. Velho, A.S.L. Ferreira, M.J. Tasso, R.S. Ferrari, R.L. Yoshida, et al., Analysis of the procedures used to evaluate suicide crime scenes in Brazil: A statistical approach to interpret reports, J. Forensic Leg. Med. **26**, 29–38, 2014. doi:10.1016/j.jflm.2014.06.004.