Evaluation of direct and indirect ethanol biomarkers using a likelihood ratio approach to identify chronic alcohol abusers for forensic purposes

Title: Evaluation of direct and indirect ethanol biomarkers using a likelihood ratio approach to identify chronic alcohol abusers for forensic purposes

Author: Eugenio Alladio Agnieszka Martyna Alberto Salomone Valentina Pirro Marco Vincenti Grzegorz Zadora

PII: S0379-0738(16)30551-5
DOI: http://dx.doi.org/doi:10.1016/j.forsciint.2016.12.019
Reference: FSI 8697

To appear in: *FSI*

Received date: 9-8-2016
Revised date: 16-11-2016
Accepted date: 12-12-2016

Please cite this article as: Eugenio Alladio, Agnieszka Martyna, Alberto Salomone, Valentina Pirro, Marco Vincenti, Grzegorz Zadora, Evaluation of direct and indirect ethanol biomarkers using a likelihood ratio approach to identify chronic alcohol abusers for forensic purposes, Forensic Science International http://dx.doi.org/10.1016/j.forsciint.2016.12.019

Evaluation of direct and indirect ethanol biomarkers using a likelihood ratio approach to identify chronic alcohol abusers for forensic purposes.

Authors:

Eugenio Alladio[1,2], Agnieszka Martyna[3], Alberto Salomone[2], Valentina Pirro[4], Marco Vincenti[1,2], Grzegorz Zadora[3,5]*

Affiliations:

[1] *Dipartimento di Chimica, Università degli Studi di Torino, Via P. Giuria 7, 10125 Torino, Italy*

[2] *Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Regione Gonzole 10/1, 10043 Orbassano, Torino, Italy.*

[3] *Department of Analytical Chemistry, Chemometric Research Group, Institute of Chemistry, The University of Silesia, Szkolna 9, 40-006 Katowice, Poland.*

[4] *Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, 47907 Indiana, USA.*

[5] *Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland.*

***corresponding author:**

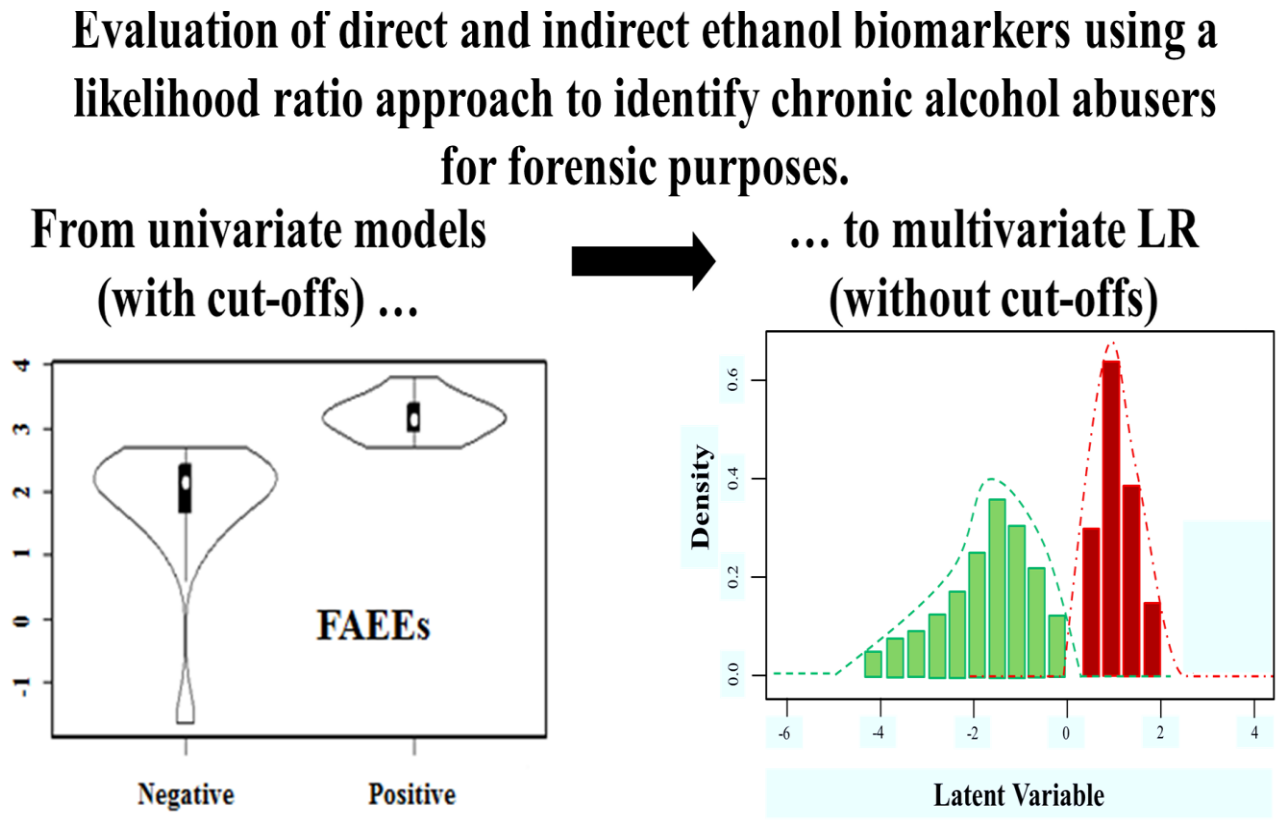
Eugenio Alladio, M.Sc.
Dipartimento di Chimica, Università degli Studi di Torino
Via Pietro Giuria 7,
10125 Torino, Italy
Tel.: +393460171979

Centro Regionale Antidoping e di Tossicologia "A. Bertinaria"
Regione Gonzole 10/1,
10043, Orbassano, Torino, Italy

E-mail: ealladio@unito.it; eugenio.alladio@gmail.com

Email of other authors:
Agnieszka Martyna, Ph.D.: ag.rzepecka@gmail.com;
Alberto Salomone, Ph.D.: alberto.salomone@antidoping.piemonte.it;
Valentina Pirro, Ph.D.: vpirro@purdue.edu;
Marco Vincenti, Prof: marco.vincenti@unito.it;
Grzegorz Zadora, Prof: gzadora@ies.krakow.pl.

Graphical abstract



Highlights

- Detection chronic alcohol misuse conditions is made possible by multivariate likelihood ratios approaches.
- Linear Discriminant Analysis in combination with likelihood ratio strategies are used to discriminate chronic from non-chronic alcohol drinkers.
- Anomalous cases related to several factors (e.g. hair treatments) can be detected, too.
- The present proof-of-concept approach might corroborate the conclusions of the traditional interpretation approach suggested by the Society of Hair Testing.
- .

Abstract

The detection of direct ethanol metabolites, such as ethyl glucuronide (EtG) and fatty acid ethyl esters (FAEEs), in scalp hair is considered the optimal strategy to effectively recognize chronic alcohol misuses by means of specific cut-offs suggested by the Society of Hair Testing. However, several factors (e.g. hair treatments) may alter the correlation between alcohol intake and biomarkers concentrations, possibly introducing bias in the interpretative process and conclusions.

125 subjects with various drinking habits were subjected to blood and hair sampling to determine indirect (e.g. CDT) and direct alcohol biomarkers. The overall data were investigated using several multivariate statistical methods. A likelihood ratio (LR) approach was used for the first time to provide predictive models for the diagnosis of alcohol abuse, based on different combinations of direct and indirect alcohol biomarkers. LR strategies provide a more robust outcome than the plain comparison with cut-off values, where tiny changes in the analytical results can lead to dramatic divergence in the way they are interpreted. An LR model combining EtG and FAEEs hair concentrations proved to discriminate non-chronic from chronic consumers with ideal correct classification rates, whereas the contribution of indirect biomarkers proved to be negligible. Optimal results were observed using a novel approach that associates LR methods with multivariate statistics. In particular, the combination of LR approach with either Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) proved successful in discriminating chronic from non-chronic alcohol drinkers. These LR models were subsequently tested on an independent dataset of 43 individuals, which confirmed their high efficiency. These models proved to be less prone to bias than EtG and FAEEs independently considered. In conclusion, LR models may represent an efficient strategy to sustain the diagnosis of chronic alcohol consumption and provide a suitable gradation to support the judgement.

Keywords

Alcohol, Likelihood ratio, Empirical cross entropy, EtG, FAEE, Hair analysis

1. Introduction

Alcohol is the most widely abused legal drug in many western countries. Health care expenditures, business and criminal justice costs associated to alcohol-related problems amount to hundreds of billions of dollars yearly, and even a greater economic burden is sustained when alcohol addictive behaviours remain untreated. Over the last decade, numerous scientific studies focused on improving the diagnosis of chronic excessive alcohol consumption to efficiently identify individuals in need of recovery programs, health care, therapeutic monitoring, etc. [1–3].

The selection of appropriate alcohol biomarkers is extremely important for correct diagnosis assessment. In fact, biased results lead to wrong analytical interpretations and consequently to clinical and/or legal errors, which can strongly impact on the life of the involved subjects. Indirect alcohol biomarkers - such as aspartate transferase (AST), alanine transferase (ALT), gamma-glutamyl transferase (GGT), mean corpuscular volume of the erythrocytes (MCV) and carbohydrate-deficient-transferrin (CDT) - measured in blood had been traditionally used to distinguish non-chronic alcohol consumers from chronic abusers [4–6]. However, they lack specificity and sensitivity [1,7–9] and have been replaced by direct alcohol biomarkers, such as ethyl glucuronide (EtG) and fatty acid ethyl esters (FAEEs), that greatly exceed indirect biomarkers in discrimination power [2,8–17]. Moreover, they are detected in the keratin matrix allowing long-term alcohol consumption monitoring.

Consensus documents of the Society of Hair Testing (SoHT) state that (i) the analysis of a 3-cm proximal scalp hair segment provides information on the average alcohol intake over a period of

about 3 months, (ii) a scalp hair concentration ≥ 30 pg/mg for EtG and ≥ 0.5 ng/mg for the sum of four FAEEs (i.e. ethyl myristate, ethyl palmitate, ethyl oleate and ethyl stearate) is indicative of chronic excessive alcohol consumption [11–13,18,19]; and (iii) the use of direct biomarkers in isolation is not advised [18]. Indeed, EtG and FAEEs absorption in hair may be altered by several factors, affecting the correlation between alcohol intake and biomarkers' concentration in hair [20,21]. For example, the hydrophilic EtG and lipophilic FAEEs have different hair incorporation mechanisms, and are differently affected by washing routines, application of alcohol-based hair care products [22], and physical-chemical hair treatments [23,24]. Therefore, their synergic use is recommended to decrease false positive rates [23–25]. Even though the consensus documents list some of the factors that may alter the analytical results and potentially introduce bias in the whole interpretative process, the interpretation of individual biomarkers results based on their respective cut-off values remains unchanged. No recommendations are given on how to interpret discordant results, nor statistical analyses are suggested to include combinations of alcohol biomarker and metadata into a predictive model.

In this study, a likelihood ratio (LR) approach is presented for the first time to better discriminate between non-chronic and chronic alcohol consumers. This approach is extensively exploited in forensics for food authentication [26,27], identification of glass [28–35], car paints [36,37], fire debris [38], inks [39], fibres [40], and DNA profiling [41,42]. LR test ($LR=Pr(E|H_1)/Pr(E|H_2)$) allows one to evaluate analytical data (E, e.g., concentrations of EtG) in the context of two mutually exclusive hypotheses (H_1: the subject is not a chronic alcohol abuser; H_2: the subject is a chronic alcohol abuser), which is what a forensic expert is asked to do in the administration of justice. More aridly, traditional interpretation models relying on cut-off values [25,40,41] are susceptible to the so-called "fall-off-cliff" problem, i.e. even minor deviations from the cut-off can utterly modify the final decision [30]. This problem is not observed when the LR test is applied because LR values not only point out which hypothesis is more consistent on the basis of the experimental evidence, but also provide the magnitude for the decision confidence thanks to the adoption of universally

accepted verbal scales [28,43] that convert LR values into statements easily comprehensible by laymen, i.e. people not expert in LR calculations.

In the present study, we tested different LR models using the scalp hair concentrations of EtG and FAEEs as experimental evidences, together with the indirect biomarkers ALT, AST, CDT, GGT, and MCV measured in whole blood. Additional investigated parameters included height, weight, and body mass index (BMI). The main goal was to investigate the discrimination power of an innovative LR approach based on multivariate statistics using different combinations of these biomarkers, in order to corroborate the diagnosis of chronic excessive alcohol consumption. The predictive capabilities of the best LR models were also tested on an independent population of 43 real caseworks individuals, including known or alleged non-chronic alcohol consumers and subjects for whom incoherent FAAEs and EtG results were determined with respect to the accepted cut-offs.

2. Materials and methods

2.1. Study protocol

The data presented herein were recovered from the databases of the Regional Antidoping and Toxicology Center "A. Bertinaria" (Orbassano , Italy). 125 subjects (118 males and 7 females) were included in this study, whose analyses were commissioned by Local Committees for Driving Licences and Alcohol Abuse Treatment Services located in Piedmont, northern Italy. Ethical approval for the study was granted by the Ethical Committee of the Azienda Ospedaliero-Universitaria San Luigi Gonzaga of Orbassano (Protocol Number 0012756). Clinical and toxicological analyses were conducted over a period of 10 months in between years 2014 and 2015. The whole blood was analyzed within 24 hours to detect ALT, AST, CDT, GGT, and MCV. Scalp hair was divided into two aliquots and measured, the proximal segment 0-3 cm was cut (no scalp hair shorter than 3 cm were analyzed), then the samples were stored at room temperature and

analyzed within 10 days to detect EtG, ethyl myristate (E14:0), ethyl palmitate (E16:0), ethyl oleate (E18:0), and ethyl stearate (E18:1). Note that within brackets are indicated the correspondent number of carbons and unsaturations (C:U) for each fatty acid. The final FAEE concentration was calculated as the sum of the four individual concentrations (i.e. E14:0, E16:0, E18:0, and E18:1). Lastly, weight and height were measured to calculate the body mass index (BMI). Only subjects under long-term monitoring at the "A. Bertinaria" Center that consistently showed negative or positive results in hair were selected to represent the population of non-chronic and chronic alcohol abusers. The archived data belonging to the individuals under examination, together with the respective clinical judgement from the medical commission in charge, allowed us to rationally divide them into the "negative" and "positive" classes, i.e. teetotallers and social drinkers (non-chronic alcohol abusers) versus chronic alcohol abusers. Subjects with doubtful classification were excluded from the study. Descriptive statistics and correlation studies were performed on the data matrix (125×12). All the analytical results are available in the Data-in-Brief [44] article associated with this study.

2.2. Determination of the direct and indirect alcohol biomarkers

Whole blood and scalp hair were collected once from each subject and analyzed within one day – for blood biomarkers – or one week – for hair biomarkers –. BD Vacutainer® EDTA and SST™ specimen tubes were used to collect whole blood samples to measure AST, ALT, GGT, CDT and MCV [45]. One of the two aliquots of hair sample was used to measure EtG. Briefly, hair samples were washed twice with methylene chloride and methanol and let to dry. Then, the samples underwent an overnight extraction step at room temperature with a 35:1 water-methanol solution, followed by sonication. Finally, approximately 100 µL of liquid phase was transferred into a vial for UHPLC–MS/MS analysis. A Shimadzu Nexera UHPLC system (Shimadzu, Duisburg,

Germany) interfaced to an AB Sciex API 5500 triple quadrupole mass spectrometer (AB Sciex, Darmstadt, Germany) was employed. EtG was detected in the negative ion mode by electrospray (ESI) ionization [46]. The second aliquot of the hair samples was used to measure FAEEs, following the same sample preparation as described in Pragst et al. [47], Suesse et al. [48] and Albermann et al.[49]. Briefly, hair samples were washed twice with n-heptane, dried at room temperature and then cut into segments (1-2 mm in length). n-heptane and DMSO were added and then samples were vortexed in a multimixer. The solvent mixture was stored at -20 °C to freeze the DMSO phase and transfer the n-heptane phase into a headspace vial. The organic solvent was dried and then reconstituted with a phosphate buffer for HS-SPME-GC/MS analysis. A MultiPurpose Sampler Flex A05-FLX-0001 (Est Analytical, West Chester Township, OH, USA) equipped with a 65 μm StableflexTM polydimethylsiloxane/divinylbenzene fiber (PDMS/DVB) from Supelco (Sigma-Aldrich, Milan, Italy) was employed in combination with a 6890N GC 5975-inert MSD (Agilent Technologies, Milan, Italy). All the methods were internally validated and accredited in accordance with ISO/IEC 17025:2005 requirements.

2.3. Data preprocessing and F-statistics

Base 10 logarithm transformation ($\log_{10}x$) was applied on each variable to reduce the skewed distribution of the data. Zero values were present in the original datasets for EtG and FAEEs. They refer to subjects whose hair samples did not yield detectable EtG and FAEE concentrations (i.e. concentration below the detection limit, LOD, of the analytical method). In order to apply the $\log_{10}$ transformation, zeros were substituted with half of the LOD value (1.5 pg/mg and 0.004 ng/mg, respectively for EtG and FAEEs. In particular, 0.004 ng/mg corresponds to half of the LOD value of ethyl palmitate, i.e. the most significant and abundant FAEE among the ones detected in this study).

An F-test feature selection procedure was used to remove the non-significant variables from the dataset, with the final intent of improving the correct classification rates and the performance of LR classification models. The F-test identifies the features that maximize the between-group variability and minimize the within-group (non-chronic vs. chronic alcohol abusers) variability [50]: the greater the F value, the better the separation between groups based on the tested variable. Only the variables with calculated F value greater than the tabulated $F_{k-1,\, l-k}$ value - where k represents the number of examined classes (k = 2) and l stands for the number of objects composing the dataset (l = 125) - at 95% significance level were considered to build LR models.

2.4. Likelihood ratio models

Different LR models were calculated and evaluated in terms of efficiency and performance. Briefly, two mutually exclusive hypotheses (H_1: the subject is not a chronic alcohol abuser – "negative" class; H_2: the subject is a chronic alcohol abuser – "positive" class) are formulated and the LR model is built with a reference population using one or more of the chosen variables (alcohol biomarkers) to represent the experimental evidence. Then, by examining the evidence for a tested subject, one of the two hypotheses is retained based on goodness of fit to either one model [28,43]. The LR value is calculated as the probability ratio that the tested subject belongs to the negative class (hypothesis H_1) or the positive class (hypothesis H_2), based on the evidence investigated. One-level models assessing only the between-object variability [26,28] were developed because the within-object variability could not be estimated, as only one measurement was completed on each individual for each parameter. Kernel density estimation (KDE) approach using Gaussian kernels was applied on the logarithmically-transformed data to estimate the between-object distributions. More details could be found in the Data-in-Brief [44] article associated with this study.

Each LR model was calculated including all 125 selected individuals, while the number of variables considered in each model was variable. Initially, twelve univariate LR models were built. Then,

different multivariate models were computed, starting with the so-called naïve LR models (LR_n) that consider all the variables simultaneously (LR_{12}) or a subset of selected variables (models LR_8, LR_7, LR_6, and LR_2) that produced significant discrimination power (measured by F-test and empirical cross entropy (ECE) values; see below). Naïve models assume that all variables are independent from one another; accordingly, they were built by multiplying the univariate LR models for the chosen variables [26,28]. Subsequently, a non-naïve multivariate LR model was evaluated using the EtG and FAEEs variables only ($LR_{FAEEs,EtG}$). Another LR model (LR_{PCA}) was calculated adopting strictly orthogonal variables after principal component analysis (PCA). Even though PCA describes a large amount of variance using few principal components, it does not necessarily mean that the corresponding information is associated with grouping. To better focus on this objective, linear discriminant analysis (LDA) was performed and a further LR model (LR_{LD}) was built on latent variables (LV). LDA evaluates the optimal direction in the PCA space that provides the best discrimination for the categories under examination and, simultaneously, estimates a certain number of delimiters according to the number of categories to be discriminated. Such direction (named LV) is represented by a linear combination of the PCs under examination. The objects are projected on the LV, converting a multivariate space into univariate, where all the individuals are represented by their projections on the LV variable. Then, the subjects are assigned to a specific category depending on their location with respect to the delimiter. In the present case, the delimiter is represented by a point along the LV direction separating the two categories (i.e. non-chronic alcohol consumers' group vs. chronic alcohol misusers' group). The LR approach was performed on this variable and a univariate LR model (LR_{LD}) was calculated. The original data were autoscaled and equal prior probabilities were adopted.

A jack-knife procedure was utilized to validate each LR model. In particular, one individual was randomly removed from the original dataset and used as a test subject to estimate the correct and false classification rates (CR), and estimate the proficiency of the developed LR model in terms of

discrimination power. Indeed, ECE values provided a quantitative measurement of the predictions strength [28,32,51], whereas CR values estimate uniquely the percentage of correct and false classifications in cross-validation without weighting the magnitude of the LR value, in other words without considering how much strong or weak is the fit of a tested object to either category [28,32]. From the exploration of ECE plots, two parameters can be extracted: $C_{llr}^{\exp}$ and C_{llr}^{cal}, indicating the reduction of information loss in terms of the amount of unexplained information [28,52] respectively for experimental and calibrated LR values. Additional information on the computation of the LR models and the ECE plots are provided in the Data-in-Brief article associated with this study [44]. All the calculations were performed with R software version 3.2.2 [53] using scripts written by the authors and the Rcmdr package [54].

3. Results and discussion

3.1. Descriptive statistics

Figure 1 shows the so-called violin plots, a combination of a box plot and a kernel density plot, for all the variables and depicts the dispersion of the logarithmically transformed data within and between the negative and positive classes. As expected, EtG and FAEEs (both the scalp hair concentration of the individual fatty acids and their sum) yielded the highest classification efficiency and their data distributions showed virtually no overlap between the two classes (Figure 1a and 1b). In particular, all negative subjects – N=96, 77% of the total cohort – showed scalp hair concentrations of both EtG and FAEEs lower than their respective cut-off values (i.e. 0.5 ng/mg and 30 pg/mg for FAEEs and EtG, respectively). Conversely, all positive individuals – N=29, 23% of the total cohort – revealed both EtG and FAEEs scalp hair concentrations higher than their respective cut-offs. These apparent over-performing results are indirectly produced by our choice of training set, including into this training set only the subjects with clear-cut drinking behaviour in

order to build robust statistical models; namely all the subjects with clinically uncertain classification were excluded. Despite this preventive selection, all indirect biomarkers display strong overlap between the distributions of the two categories. Figure 1c and 1d show the distributions observed for CTD and GGT as examples. Although the data appear not to have Gaussian distributions, LR models are appropriate, because they properly work also with non-parametric distributions.

3.2. Univariate LR models

Univariate LR models, together with their respective ECE plots, were evaluated for each of the twelve variables. The CR responses and the $C_{llr}^{\exp}$ and C_{llr}^{cal} values for univariate LR models using the KDE approach are reported in Table 1. As expected, univariate LR models for FAEEs and EtG provided the best global correct classification rate (about 96% and 95%, respectively). Their ECE plots showed satisfactory results in terms of accuracy, calibration and discrimination power (Figure 2a-2b). On the other hand, indirect biomarkers provided poor CR and $C_{llr}^{\exp}$ results, with the worst performance observed for MCV ($C_{llr}^{\exp}$ = 116%), meaning that their single value delivers misleading information, as already reported in other scientific studies [1,9,13,45]. ECE plots relevant to the univariate LR models of all the indirect and E14:0, E16:0, E18:0 and E18:1 direct biomarkers are available in the Data-in-Brief article associated to the present one [44]. For LR models of FAEEs and EtG, the information loss was still accountable, with $C_{llr}^{\exp}$ values equal to 24% and 22%, respectively, where zero represents the ideal value for $C_{llr}^{\exp}$ and C_{llr}^{cal}, with no information loss and systematic support to the correct hypothesis from the evidence.

3.3. Multivariate LR models

3.3.1. Naïve multivariate LR model

The correlation coefficients matrix reported in Table 2 indicated stronger correlations between the individual FAEEs and their sum, and between EtG and FAEEs, as expected. Notably, FAEEs and E16:0 show the highest correlation coefficient (i.e. 0.96), in agreement with the proposed update of SoHT consensus documents, where the single ethyl palmitate is proposed to substitute the sum of four FAEEs in the interpretation criteria. Although the naïve approach theoretically requires variables with no significant correlations, an eligible naïve multivariate LR model (LR_{12}) was developed taking into account all the variables together. Singularly, it was possible to accept the lack of correlation assumption since the number of individuals constituting the database was limited.

The naïve LR_{12} model provided better CR value (98.4%) than the univariate methods for FAEEs and EtG. Only two negative individuals (i.e. non-chronic alcohol abusers) out of 96 were misclassified as positive subjects, leading to a correct classification rate for the negative class of 97.9% (Table 1). Accuracy and calibration proved satisfactory, as well as the reduction of information loss with $C_{llr}^{\exp}$ and C_{llr}^{cal} values equal to 15.6% and 4.1%, respectively (Table 1). However, further LR models were investigated with the aim of selecting a lower number of variables and decreasing the amount of redundant information and noise.

3.3.2. Naïve and non-naïve multivariate LR model relevant to the variables selected by F-test and ECE plots

Further LR models were built using the selected variables that showed statistically significant discriminant power ($F_{\alpha=0.05,1,123} > 3.92$). The ECE curve shape was also used as a feature selection criterion, i.e. the variables whose experimental LR values, represented by the solid red line, exceeded the null curve, represented by the dotted black line, were excluded (Figures 2 and 3). A new naïve model based on CDT, GGT, E14:0, E16:0, E18:0, E18:1, FAEEs, and EtG was built (LR_8). This new naïve model showed the same CR as the naïve LR_{12} model (98.4%) and the same

number of misleading classifications; however, the ECE plot showed improvement in the reduction of information loss, with $C_{llr}^{\exp}$ and C_{llr}^{cal} values lowered to 4.4% and 1.7%. The analysis was repeated after taking out GGT from the model (LR_7) and then excluding also CDT (LR_6), as they had the lowest significant discrimination power among the eight variables. No changes were observed in the correct classification rates nor in the information loss (Table 1). To stretch the system even further, a naïve multivariate model with only the FAEEs and EtG variables was built (LR_2, Figure 3a). In general, FAEEs and EtG proved to provide the best discrimination between non-chronic and chronic alcohol abusers, while the contribution of indirect biomarkers is negligible, as several other independent studies concluded [1,3,8,9,13,45].

A non-naïve multivariate model with FAEEs and EtG variables was computed ($LR_{FAEEs,EtG}$). Improved results were obtained in comparison to LR_2 and LR_6 in terms of $C_{llr}^{\exp}$, C_{llr}^{cal} and ECE values (Table 1 and Figure 3b) suggesting that the correlation between the two biomarkers carries useful information for the decision-making process. Further LR models were tested to investigate whether the multivariate evaluation of all the different FAEE biomarkers (i.e. E14:0, E16:0, E18:0, E18:1) might provide better performance than the LR model investigating the sum of the four FAEE concentrations (i.e. FAEEs). The univariate LR model involving FAEEs only (LR_{FAEEs}) provided unsatisfactory $C_{llr}^{\exp}$ and C_{llr}^{cal} values. On the other hand, both naïve and non-naïve multivariate LR models ($LR_{m,naïve}$; $LR_{m,non\text{-}naïve}$) showed unchanged CR rates, but lower $C_{llr}^{\exp}$ and C_{llr}^{cal} values (Table 1). Thus, the multivariate approach is apparently preferable to the current interpretation approach that inspects the sum of FAEEs at the univariate level. However, further confirmations are required on a larger population in order to better compare the two approaches.

3.3.3. Multivariate LR model based on the variables selected by F-test and ECE plots analysis and orthogonalised by PCA

In order to improve further the LR model, the variables selected by F-tests and ECE plots analysis (i.e., E14:0, E16:0, E18:0, E18:1, FAEEs and EtG) were orthogonalized by means of the PCA technique. The new multivariate model (LR_{PCA}) was built on the first three principal components (describing more than 95% of cumulative variance), and computed by multiplying the LR values from univariate LR frameworks relative to the selected principal components. A satisfactory CR equal to 99.2% was observed, but no improvement of ECE plot (Figure 3e) and $C_{llr}^{\exp}$ value (6.5%) was observed (C_{llr}^{cal} = 1.7%). The lower reduction of information loss observed for LR_{PCA} with respect to the previous naïve multivariate LR models depends on the strong correlation occurring among the selected variables that compose the PCs.

3.3.4. Likelihood ratio model based on linear discriminant analysis

Linear discriminant analysis (LDA) is nowadays widely used in several fields of chemistry [26,27,55,56]. The LDA technique was exploited with the aim of identifying the direction within the PCA space that better discriminate non-chronic from chronic alcohol drinkers (see Materials and Methods). The projection of the objects from the PCA space onto the LV variable produced suitable separation between the two categories, as shown in Figure 4, in which the continuous and dashed lines represent the kernel density plot for the non-chronic and chronic alcohol consumers, respectively. The correct classification rates and the ECE curve parameters of the LR_{LD} model reported in Table 1 and Figure 3f exhibit optimal results in terms of correct classification rates (i.e., overall CR equal to 100%) and reduction of information loss ($C_{llr}^{\exp}$ = 3.3%; C_{llr}^{cal} = 0.0%), outscoring all the previous LR models. In conclusion, all the multivariate LR models provided better

performances than the univariate LR models (Table 1), but the LR_{LD} model produced the best outcomes in terms of accuracy and discriminating power.

3.4. Real caseworks

The best performing non-naïve ($LR_{FAEEs,EtG}$) and LR_{LD} models were used to evaluate 43 new individuals (Table 3), consisting of 6 known non-chronic alcohol consumers (case1-6), 5 individuals (case7-11) expected to be chronic alcohol consumers (according to their historical data, which showed a consistent number of analytical results above SoHT cut-off values over the time), one individual showing extreme and conflicting results from the direct biomarkers (case12), and 31 selected individuals with unknown disposition towards alcohol consumption, but whose EtG and FAEEs scalp hair concentrations were conflicting and incoherent with respect to the SoHT cut-off valus (case13-43). The experimental EtG and FAEE data, together with LR values and support to the most probable hypothesis, are reported in Table 3. In particular, the verbal scale that was used to convert the LR numerical values into different confidence expressions in support to a certain conclusion is reported in Table 4, according to the literature [28,43]. The negative case1-case6 were correctly identified as non-chronic alcohol consumers with LR values largely above unity ($2.1 \cdot 10^{3}$-$7.4 \cdot 10^{9}$) from both models. For cases7-case11, all individuals were correctly identified as positive (i.e. chronic alcohol abusers) with LR values ranging from $4.1 \cdot 10^{-2}$ to $2.8 \cdot 10^{-4}$ for both models.

Case12 was classified as negative, with very strong support for non-chronic hypothesis H_1 ($LR_{FAEEs,EtG}$ value equal to $1.3 \cdot 10^{64}$). His scalp hair specimen had an extremely high concentration of EtG (2769 pg/mg) but an extremely low concentration of FAEEs (approx. 0.01 ng/mg). Further investigation on this case pointed out that the subject used to apply an Arnica-based oil lotion on his hair that contained EtG. Interestingly, the $LR_{FAEEs,EtG}$ model did not classify this subject as a chronic alcohol abuser, as the EtG concentration alone would suggest, which strongly highlights the robustness of the LR approach. In fact, LR takes into account the rarity of the measured analytical data, whereas this is typically ignored by other non-Bayesian discriminant methodologies (both

univariate and multivariate) even if strongly significant, especially in forensic caseworks. Bayesian discriminant methods typically deliver strong support to one of the hypotheses when rare physicochemical data are compared with a the reference "normal" population. On the other hand, the LR_{LD} model classified case12 as positive, with quite strong support for H_2 (1/LR value equal to 6250). However, this individual could be easily recognized as an outlier from Hotelling T^2 scores and Q residuals [57]. In practice, the Hotelling T^2 vs. Q residuals plot is employed in PCA to recognize samples that present very rare features, and differentiate them from the reference populations. For this reason, case12 could be identified as anomalous anyway, even though an external factor (i.e., the use of an Arnica-based oil lotion containing EtG) produced a bias on LDA features and results.

Predictions for the remaining 31 cases (Table 3) varied from "negative with very strong support" to "positive with strong support" (i.e., case13). Fourteen cases (from case14 to case27) are particularly interesting and represent a variety of real situations of difficult judgement, where either one of biomarkers largely exceeds the cut-off while the other is far below, or both are very close to the corresponding cut-off values. The LR approach not only helps the toxicologist solve these puzzles, but also provide a quantitative support to his/her decision. For example, case18 presented an extremely high concentration of FAEE (11.98 ng/mg) together with a 26 pg/mg EtG concentration value, slightly lower than the corresponding cut-off of 30 pg/mg. The SoHT consensus document (18) indicates EtG as the deciding biomarker in the cases of ambiguous results with respect to FAEEs data. In a stringent interpretation of the consensus document, case18 should have been classified as non-chronic despite the huge FAEEs concentration observed and the EtG concentration close to cut-off. In contrast, both LR approaches provided a moderately strong support to the chronic alcohol misuse hypothesis H_2 ($LR_{FAEEs,EtG}$ value equal to $5.3 \cdot 10^{-3}$, LR_{LD} value equal to $1.9 \cdot 10^{-3}$), which appears. more reliable than H_1, according to the experimental results. Remarkably, all 31 individuals with unknown disposition towards alcohol were classified with identical response,

negative or positive, by both $LR_{FAEEs,EtG}$ and LR_{LD} models, which also provided similar strengths in supporting either one hypothesis according to EtG and FAEE values (Table 3). This can be also observed in the correlation plot (Figure 5), where the comparison of the Log_{10}-LR values provided a highly significant coefficient of determination (R^2=0.9614). This result suggests that the models behave similarly in interpreting FAEE and EtG data, and classify the unknown individuals with the same final judgement. The simultaneous use of two LR models is likely to represent a powerful interpretation approach in order to solve the ambiguous caseworks where EtG and FAEEs hair concentrations turn to be incoherent with respect to the SoHT cut-offs.

4. Conclusions

For the identification of chronic excessive alcohol consumption, SoHT presently suggests a cut-off based interpretation model established on EtG and FAEE levels detected in scalp hair and hair sampled from other body sites with the exception of axillary and pubic hair regions. However, the concentration of EtG and FAEEs in hair can be influenced by several factors (e.g. the use cosmetic treatments and thermal hair straightening tools), occasionally leading to challenging interpretation when the classical univariate approach is adopted, since minor data changes may completely reverse the final decision. The present study goes beyond the cut-off based decision method and proposes to exploit the advantages arising from the combined use of LR approach and multivariate statistics for the interpretation of FAEE and EtG values in hair samples. The LR approach proved to represent a discriminant strategy that does not rely on a fixed threshold value to make predictions but rather relies on probability distributions. The most remarkable advantage offered by LR models is that they provide different levels of strength in supporting a hypothesis on the basis of the experimental data. Furthermore, LR approaches take into account the information about the rarity of the physicochemical data, allowing the identification of anomalous values that might have been influenced by external factors (e.g. cosmetic habits), which are not commonly represented in the

reference population utilized to build the LR model. On the other hand, multivariate statistics add value to the information arising from the single biomarkers, whose combined interpretation decreases the probability of false positive and false negative outcomes. Several naïve and non-naïve multivariate LR models have been tested in the present study, suggesting powerful alternatives to the classical univariate interpretation approach. In particular, two LR models provided suitable discrimination between non-excessive alcohol consumers and chronic alcohol abusers that respectively (i) evaluate EtG and FAEE variables in a non-naïve multivariate model ($LR_{FAEEs,EtG}$), and (ii) combine linear discriminant analysis with the LR approach (LR_{LD}). Both models provided high rates of correct classification, satisfactory ECE curve parameters, similar outcomes and powerful prediction capabilities. If employed together, $LR_{FAEEs,EtG}$ and LR_{LD} models allowed to efficiently interpret EtG and FAEEs scalp hair concentrations, even if incoherent with respect to the SoHT cut-off values and conflicting. Furthermore, their simultaneous use allowed to recognize and interpret the anomalous values related to the influence of endogenous and/or external factors. At the current stage, LR multivariate models represent a supporting tool whose outcomes are still continuously tested, but the adoption of LR strategies appears to provide remarkable results in terms of robustness and discrimination power with respect to the classical univariate approach. Further validation of this LR interpretation approach will arise from its introduction into the routine workflow of our forensic toxicology laboratory, in order to support the diagnosis of chronic alcohol consumption and assist forensic analysts in the decision-making process.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

No specific financial support was received for this study. Continuous support from M.I.U.R. and Regione Piemonte is kindly acknowledged.

References

[1] R. Boscolo-Berto, D. Favretto, G. Cecchetto, M. Vincenti, R. Kronstrand, S.D. Ferrara, G. Viel, Sensitivity and specificity of EtG in hair as a marker of chronic excessive drinking: pooled analysis of raw data and meta-analysis of diagnostic accuracy studies., Ther Drug Monit. 36 (2014) 560-575.

[2] M. Hastedt, S. Herre, F. Pragst, M. Rothe, S. Hartwig, Workplace alcohol testing program by combined use of ethyl glucuronide and fatty acid ethyl esters in hair, Alcohol Alcohol. 47 (2012) 127–132.

[3] M. Hastedt, M. Büchner, M. Rothe, R. Gapert, S. Herre, F. Krumbiegel, M. Tsokos, T. Kienast, A. Heinz, S. Hartwig, Detecting alcohol abuse: Traditional blood alcohol markers compared to ethyl glucuronide (EtG) and fatty acid ethyl esters (FAEEs) measurement in hair, Forensic Sci. Med. Pathol. 9 (2013) 471–477.

[4] P.R. Marques, Levels and types of alcohol biomarkers in DUI and clinic samples for estimating workplace alcohol problems., Drug Test. Anal. 4 (2012) 76–82.

[5] F. Pragst, Chapter 4 – Alcohol Biomarkers in Hair, in Hair Analysis in Clinical and Forensic Toxicology, 1st Ed. Elsevier, San Diego, CA, USA, 2015; 71-139.

[6] M.L. Hannuksela, M.K. Liisanantti, A.E. Nissinen, M.J. Savolainen, Biochemical markers of alcoholism, Clin. Chem. Lab. Med. 45 (2007) 953–961.

[7] R. Agius, T. Nadulski, H.G. Kahl, B. Dufaux, Ethyl glucuronide in hair - A highly effective test for the monitoring of alcohol consumption, Forensic Sci. Int. 218 (2012) 10–14.

[8] C.L. Crunelle, M. Yegles, A.L.N. Van Nuijs, A. Covaci, M. De Doncker, K.E. Maudens, B. Sabbe, G. Dom, W.E. Lambert, P. Michielsen, H. Neels, Hair ethyl glucuronide levels as a marker for alcohol use and abuse: A review of the current state of the art, Drug Alcohol Depend. 134 (2014) 1–11.

[9] V. Pirro, V. Valente, P. Oliveri, A. De Bernardis, A. Salomone, M. Vincenti, Chemometric evaluation of nine alcohol biomarkers in a large population of clinically-classified subjects: pre-eminence of ethyl glucuronide concentration in hair for confirmatory classification., Anal. Bioanal. Chem. 401 (2011) 2153–2164.

[10] L. Morini, C. Varango, C. Filippi, C. Rusca, P. Danesino, F. Cheli, M. Fusini, G. Iannello, A. Groppi, Chronic excessive alcohol consumption diagnosis: comparison between traditional biomarkers and ethyl glucuronide in hair, a study on a real population., Ther. Drug Monit. 33 (2011) 654–657.

[11] P.R. Marques, A.S. Tippetts, M. Yegles, Ethylglucuronide in hair is a top predictor of impaired driving recidivism, alcohol dependence, and a key marker of the highest BAC interlock tests., Traffic Inj. Prev. 15 (2014) 361–369.

[12] V. Kulaga, J. Gareri, N. Fulga, G. Koren, Agreement between the fatty acid ethyl ester hair test for alcohol and social workers' reports., Ther. Drug Monit. 32 (2010) 294–299.

[13] H. Kharbouche, M. Faouzi, N. Sanchez, J.B. Daeppen, M. Augsburger, P. Mangin, C. Staub, F. Sporkert, Diagnostic performance of ethyl glucuronide in hair for the investigation of alcohol drinking behavior: a comparison with traditional biomarkers., Int. J. Legal Med. 126 (2012) 243–250.

[14] S.H. Stewart, D.G. Koch, I.R. Willner, P.K. Randall, A. Reuben, Hair ethyl glucuronide is highly sensitive and specific for detecting moderate-to-heavy drinking in patients with liver disease., Alcohol Alcohol. 48 (2013) 83–87.

[15] P. Kintz, D. Nicholson, Testing for ethanol markers in hair: Discrepancies after simultaneous quantification of ethyl glucuronide and fatty acid ethyl esters., Forensic Sci. Int. 243 (2014) 44–46.

[16] P. Kintz, Toxicological aspects of drug-facilitated crimes, Acad. Press. 4 (2014) 102–107.

[17] M.E. Albermann, F. Musshoff, B. Madea, Comparison of ethyl glucuronide (EtG) and fatty acid ethyl

esters (FAEEs) concentrations in hair for testing abstinence., Anal. Bioanal. Chem. 400 (2011) 175–181.

[18] Society of Hair Testing, Consensus for the use of alcohol markers in hair for assessment of both abstinence and chronic excessive alcohol consumption, Bordeaux, 2014.

[19] L. Morini, L. Politi, S. Acito, A. Groppi, A. Polettini, Comparison of ethyl glucuronide in hair with carbohydrate-deficient transferrin in serum as markers of chronic high levels of alcohol consumption., Forensic Sci. Int. 188 (2009) 140–143.

[20] M. Vincenti, P. Kintz. Chapter 12 – New Challenges and Perspectives in Hair Analysis in Hair Analysis in Clinical and Forensic Toxicology, 1st Ed. Elsevier, San Diego, CA, USA, 2015; 337–368.

[21] A. Salomone, V. Pirro, T. Lombardo, D. Di Corcia, S. Pellegrino, M. Vincenti, Interpretation of group-level factors from a large population dataset in the determination of ethyl glucuronide in hair., Drug Test. Anal. (2014) 407–413.

[22] F. Sporkert, H. Kharbouche, M.P. Augsburger, C. Klemm, M.R. Baumgartner, Positive EtG findings in hair as a result of a cosmetic treatment, Forensic Sci. Int. 218 (2012) 97–100.

[23] F. Pragst, M. Yegles, Determination of fatty acid ethyl esters (FAEE) and ethyl glucuronide (EtG) in hair: a promising way for retrospective detection of alcohol abuse during pregnancy?, Ther. Drug Monit. 30 (2008) 255–263.

[24] S. Suesse, F. Pragst, T. Mieczkowski, C.M. Selavka, A. Elian, H. Sachs, M. Hastedt, M. Rothe, J. Campbell, Practical experiences in application of hair fatty acid ethyl esters and ethyl glucuronide for detection of chronic alcohol abuse in forensic cases, Forensic Sci. Int. 218 (2012) 82–91.

[25] F. Pragst, M. Rothe, B. Moench, M. Hastedt, S. Herre, D. Simmert, Combined use of fatty acid ethyl esters and ethyl glucuronide in hair for diagnosis of alcohol abuse: Interpretation and advantages, Forensic Sci. Int. 196 (2010) 101–110.

[26] A. Martyna, G. Zadora, I. Stanimirova, D. Ramos, Wine authenticity verification as a forensic problem: An application of likelihood ratio test to label verification, Food Chem. 150 (2014) 287–

295.

[27] P. Własiuk, A. Martyna, G. Zadora, A likelihood ratio model for the determination of the geographical origin of olive oil, Anal. Chim. Acta. 853 (2015) 187–199.

[28] G. Zadora, A. Martyna, D. Ramos, C. Aitken. Statistical Analysis in Forensic Science - Evidential Value of Multivariate Physicochemical Data. First ed. Chicester, UK: John Wiley & Sons, Ltd.; 2014. 1-322.

[29] C.G.G. Aitken, D. Lucy, Evaluation of trace evidence in the form of multivariate data, Appl. Stat. 53 (2004) 109–122.

[30] C.G.G. Aitken, G. Zadora, D. Lucy, A Two-Level Model for Evidence Evaluation, J. Forensic Sci. 52 (2007) 412–419.

[31] A. Martyna, K.E. Sjastad, G. Zadora, D. Ramos, Analysis of lead isotopic ratios of glass objects with the aim of comparing them for forensic purposes, Talanta. 105 (2013) 158–166.

[32] D. Ramos, G. Zadora, Information-theoretical feature selection using data obtained by scanning electron microscopy coupled with and energy dispersive X-ray spectrometer for the classification of glass traces., Anal. Chim. Acta. 705 (2011) 207–217.

[33] G. Zadora, Classification of Glass Fragments Based on Elemental Composition and Refractive Index, J. Forensic Sci. 54 (2009) 49–59.

[34] G. Zadora, T. Neocleous, Evidential Value of Physicochemical Data - Comparison of Methods of Glass Database Creation, J. Chemom. 24 (2010) 367–378.

[35] G. Zadora, D. Ramos, Evaluation of glass samples for forensic purposes — An application of likelihood ratios and an information–theoretical approach, Chemometr. Intell. Lab. 102 (2010) 63–83.

[36] A. Martyna, A. Michalska, G. Zadora, Interpretation of FTIR spectra of polymers and Raman spectra of car paints by means of likelihood ratio approach supported by wavelet transform for reducing data dimensionality, Anal. Bioanal. Chem. 407 (2015) 3357–3376.

[37] A. Michalska, A. Martyna, J. Zieba-Palus, G. Zadora, Application of likelihood ratio approach in

solving a comparison problem of Raman spectra recorded for blue automotive, J. Raman Spectrosc. 46 (2015) 772–783.

[38] G. Zadora, R. Borusiewicz, J. Zieba-Palus, Differentiation between weathered kerosene and diesel fuel using automatic thermal desorption-GC-MS analysis and the likelihood ratio approach, J. Sep. Sci. 28 (2005) 1467–1475.

[39] A. Martyna, D. Lucy, B.M. Trzcinska, D. Ramos, A. Parczewski, The evidential value of microspectrophotometry measurements made for pen inks, Anal. Methods. 5 (2013) 6788–6795.

[40] G. Zadora, Evaluation of evidential value of physicochemical data by a Basyesian network approach, J. Chemom. 24 (2010) 346–366.

[41] P. Gill, C.H. Brenner, J.S. Buckleton, a. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir, DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures, Forensic Sci. Int. 160 (2006) 90–101.

[42] R. Puch-Solis, P. Roberts, S. Pope, C.G.G. Aitken, Assessing the Probative Value of DNA Evidence: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnessess. Practitioner Guide No. 2, London: Royal Statistical Society, 2013.

[43] C.G.G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd Edition, John Wiley & Sons, Inc., 2004.

[44] E. Alladio, A. Martyna, A. Salomone, V. Pirro, M. Vincenti, Zadora G, Multivariate data analysis and likelihood ratio for the interpretation of direct and indirect alcohol biomarkers in hair samples, Forensic Sci. Int. Data-in-Brief (submitted).

[45] V. Pirro, P. Oliveri, B. Sciutteri, R. Salvo, A. Salomone, S. Lanteri, M. Vincenti, Multivariate strategies for screening evaluation of harmful drinking, Bioanal. 5 (2013) 687–699.

[46] V. Pirro, D. Di Corcia, F. Seganti, A. Salomone, M. Vincenti, Determination of ethyl glucuronide levels in hair for the assessment of alcohol abstinence, Forensic Sci. Int. 232 (2013) 229–236.

[47] F. Pragst, V. Auwaerter, F. Sporkert, K. Spiegel, Analysis of fatty acid ethyl esters in hair as possible

markers of chronically elevated alcohol consumption by headspace solid-phase microextraction (HS-SPME) and gas chromatography-mass spectrometry (GC-MS), Forensic Sci. Int. 121 (2001) 76–88.

[48] S. Suesse, C.M. Selavka, T. Mieczkowski, F. Pragst, Fatty acid ethyl ester concentrations in hair and self-reported alcohol consumption in 644 cases from different origin, Forensic Sci. Int. 196 (2010) 111–117.

[49] M.E. Albermann, B. Madea, F. Musshoff, A SPME-GC/MS procedure for the determination of fatty acid ethyl esters in hair for confirmation of abstinence test results, J. Chromatogr. Sci. 218 (2013) 10–14.

[50] M. Otto, Chemometrics: Statistics and computer application in analytical chemistry., 2nd ed., Weinheim, Germany, 2007.

[51] N. Brümmer, N. du Preez, Application-independent evaluation of speaker detection, Comput. Speech Lang. 20 (2006) 230–275.

[52] P. Wlasiuk, A. Martyna, G. Zadora, An effect of selection of a smoothing parameter on the performance of likelihood ratio models for solving classi fi cation task, 144 (2015) 87–99.

[53] R Core Team, R: A language and environment for statistical computing, Vienna, Austria, 2015.

[54] J. Fox, The R Commander: A Basic Statistics Graphical User Interface to R, J. Stat. Softw. 14 (2005) 1–42.

[55] R. Leardi, Chemometrics: from classical to genetic algorithms, Grasas Y Aceites. 53 (2002) 115–127.

[56] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, First ed., Amsterdam, 1997.

[57] J. Kuligowski, G. Quintás, C. Herwig, B. Lendl, A rapid method for the differentiation of yeast cells grown under carbon and nitrogen-limited conditions by means of partial least squares discriminant analysis employing infrared micro-spectroscopic data of entire yeast cells., Talanta. 99 (2012) 566–73.

Figures

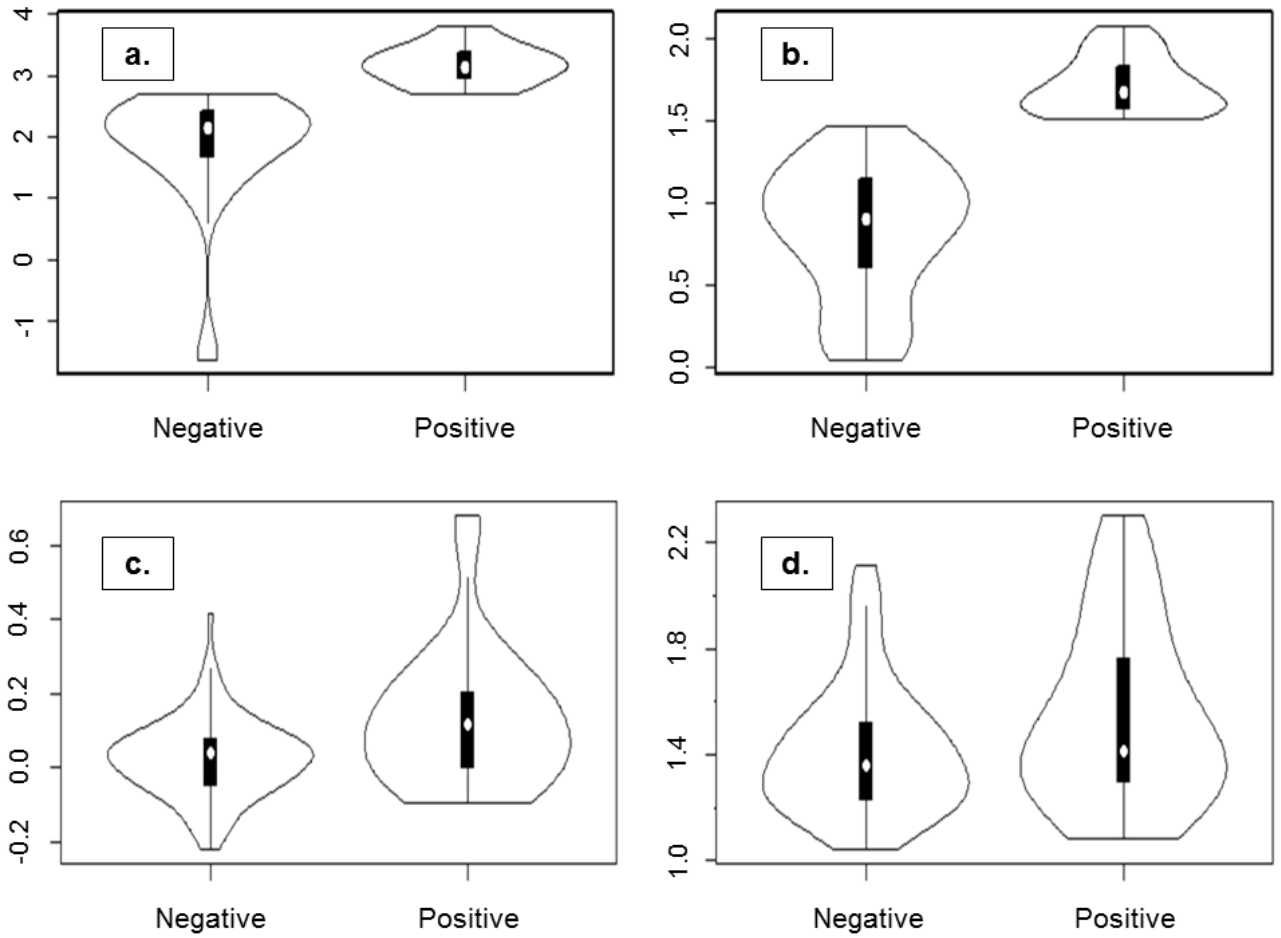


Figure 1 Violin plots (i.e. a combination of box plots and a kernel density estimation plots) relative to FAEEs (a), EtG (b), CDT (c) and GGT (d) variables. Individuals were divided into two categories, where "Negative" represents the group of non-chronic alcohol consumers and "Positive" stands for the group of chronic alcohol misusers

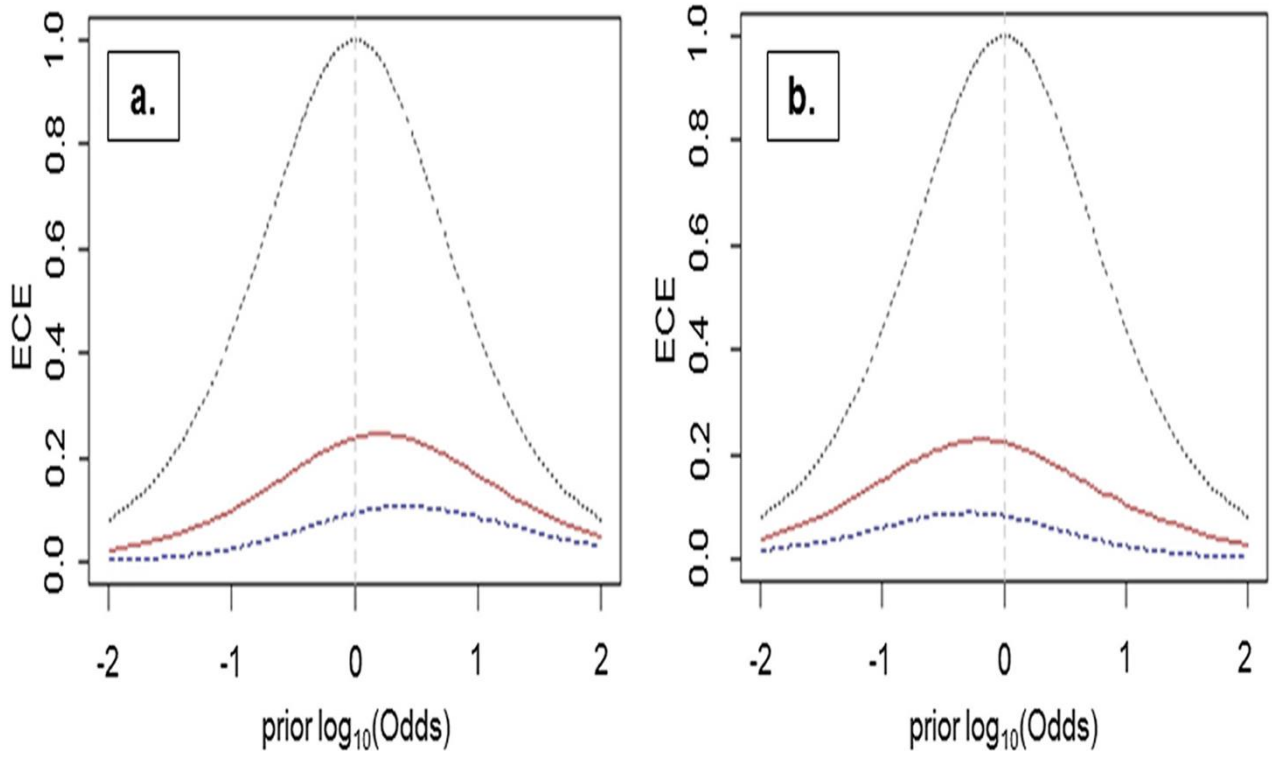


Figure 2 The ECE plots describing the performance of univariate LR models relevant to FAEEs (a), EtG (b) variables

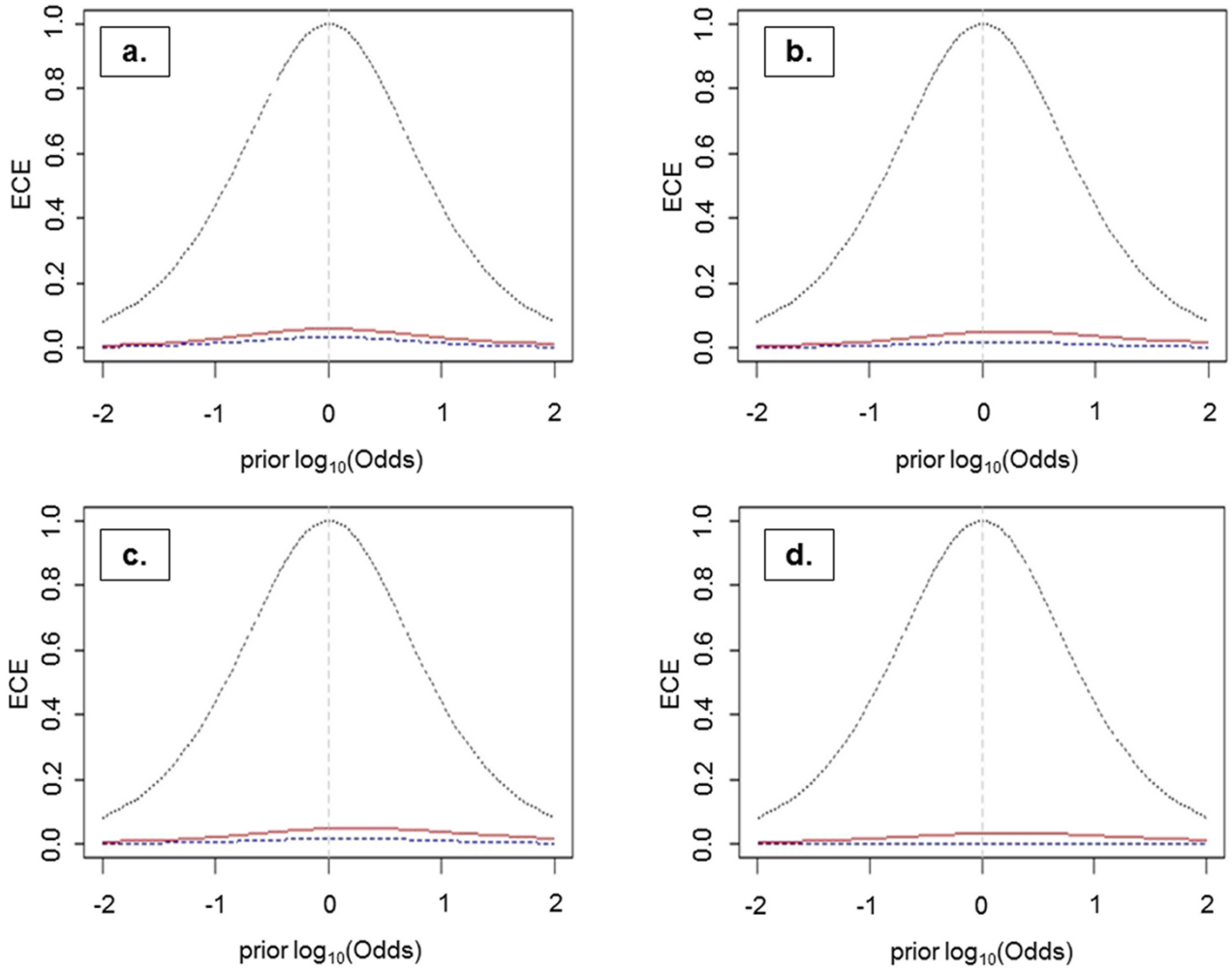


Figure 3 ECE plots relevant to the developed multivariate naïve LR models involving: (a) FAEEs and EtG variables only (LR_2); (b) FAEEs and EtG variables only employing a non-naïve multivariate LR model ($LR_{FAEEs,EtG}$); (c) the features selected by F-test and ECE plots analysis and orthogonalized by PCA (LR_{PCA}); (d) the variable (named LD) from the LDA approach representing a linear combination of the features that were selected by means of F-test and ECE plots analysis (LR_{LD})

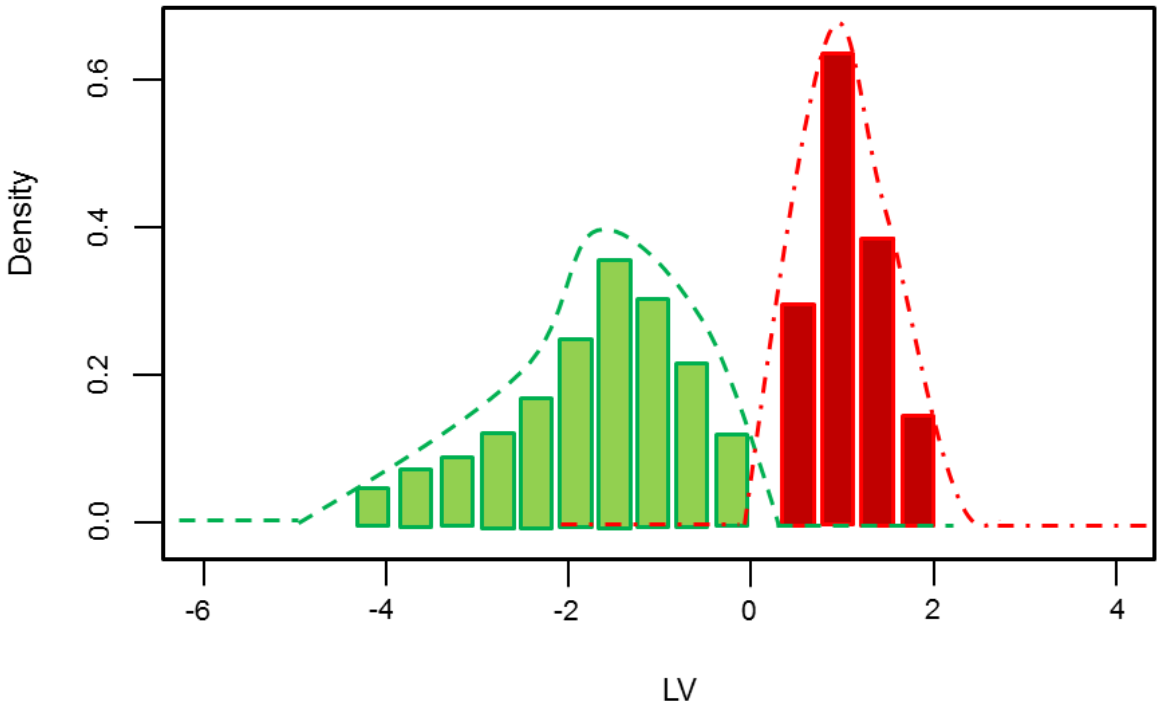


Figure 4 Comparative Kernel Density Plot relevant to the LV values of the non-chronic and chronic alcohol consumers. In particular, a green solid line represents the individuals belonging to the group of the non-chronic alcohol consumers ("Negative"), while a red dashed line describes the individuals belonging to the group of the chronic alcohol misusers ("Positive")

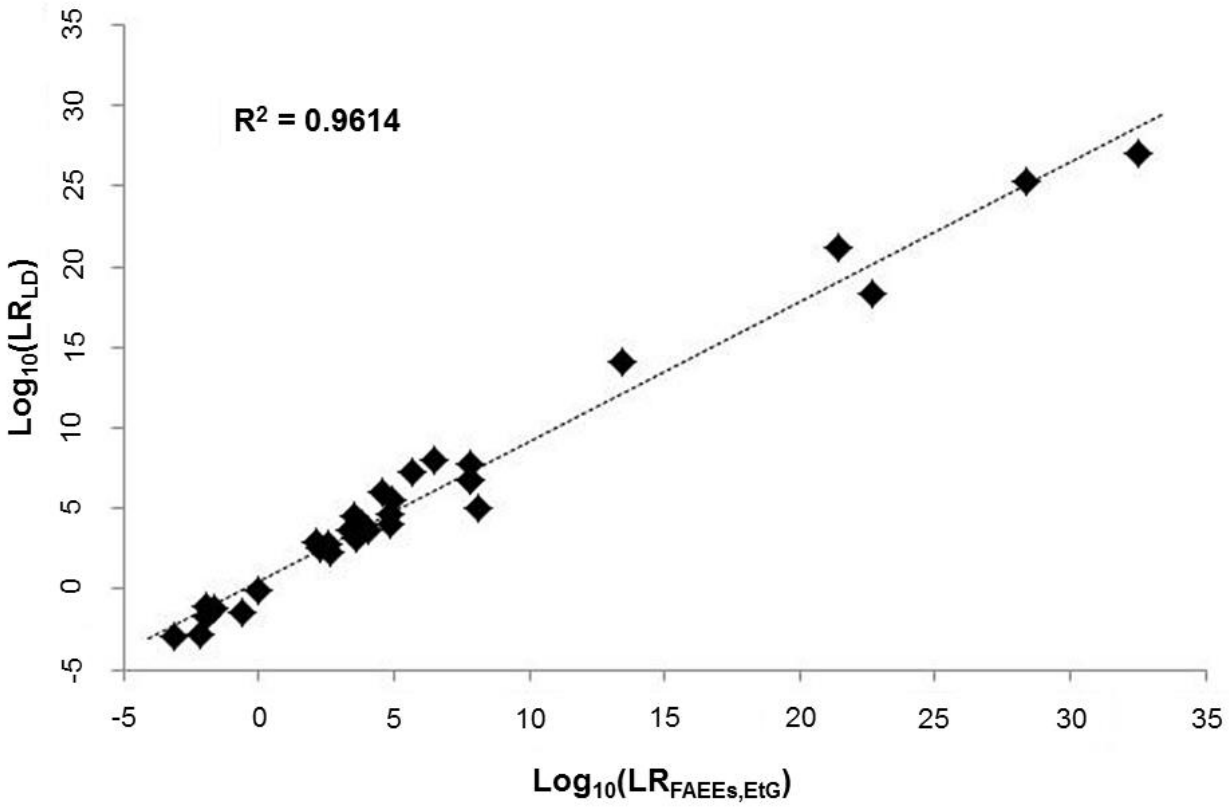


Figure 5 Comparison of Log_{10}-LR values provided by $LR_{FAEEs,EtG}$ (x) and LR_{LD} (y) models. The dashed line represents the ideal situation where the two LR model provide the same LR value. A significant coefficient of determination (R^2) equal to 0.9614 is observed.

Tables

Table 1 The rates of correct classification within each classification problem (CR [%]), as the weighted sum of the rates of correct classification for the two categories under examination (CR_Neg [%], CR_Pos [%]; CR_Neg – Negative, non-chronic alcohol consumers' group; CR_Pos – Positive, chronic alcohol consumers' group). The ECE parameters describing the performance of the LR models are reported too, by means of the values for $\log_{10}$(prior odds)=0 (i.e. $C_{llr}^{\exp}$ [%] and C_{llr}^{cal} [%] values for the experimental and calibrated curves, respectively)

Univariate LR models					
LR model	**CR**	**CR_Neg**	**CR_Pos**	$C_{llr}^{\exp}$	C_{llr}^{cal}
ALT	66.4	75.0	37.9	102.1	95.5
AST	56.0	67.7	17.2	102.4	97.3
CDT	76.8	86.5	44.8	95.7	88.1
E14:0	88.0	86.5	93.1	46.4	35.4
E16:0	94.4	93.8	96.6	28.3	16.5
E18:0	88.8	88.5	89.7	39.6	28.5
E18:1	92.0	89.6	100.0	26.9	12.7
EtG	95.2	94.8	96.6	22.3	8.2
FAEEs	96.0	94.8	100.0	23.7	9.4
GGT	70.4	81.2	34.5	99.3	94.5
MCV	71.2	79.2	44.8	116.2	83.7

Multivariate LR models						
LR model	**CR**	**CR_Neg**	**CR_Pos**	$C_{llr}^{\exp}$	C_{llr}^{cal}	**Variables**
LR_{12}	98.4	97.9	100.0	15.6	4.1	All the variables
LR_{8}	98.4	97.9	100.0	4.4	1.7	E14:0,E16:0,E18:0,E18:1,ETG,FAEEs,CDT,GGT
LR_{7}	98.4	97.9	100.0	4.5	1.7	E14:0,E16:0,E18:0,E18:1,ETG,FAEEs,CDT
LR_{6}	98.4	97.9	100.0	5.2	2.5	E14:0,E16:0,E18:0,E18:1,ETG,FAEEs
LR_{2}	98.4	99.0	96.6	5.9	3.2	ETG,FAEEs
$LR_{FAEEs,EtG}$	98.4	99.0	96.6	4.9	1.7	ETG,FAEEs
$LR_{m,naïve}$	96.0	94.8	100.0	21.1	8.18	E14:0,E16:0,E18:0,E18:1
$LR_{m,non-naïve}$	96.0	94.8	100.0	21.1	8.18	E14:0,E16:0,E18:0,E18:1
LR_{PCA}	99.2	97.9	100.0	6.5	1.7	E14:0,E16:0,E18:0,E18:1,ETG,FAEEs
LR_{LD}	100.0	100.0	100.0	3.3	0.0	E14:0,E16:0,E18:0,E18:1,ETG,FAEEs

Table 2 Partial correlation coefficients matrix for the 12 variables collected. The upper part of the matrix is omitted as it is symmetric to the lower one

	E14:0	**E16:0**	**E18:1**	**E18:0**	**FAEEs**	**ETG**	**AST**	**ALT**	**GGT**	**MCV**	**CDT**	**BMI**
E14:0	1											
E16:0	0.7	1										
E18:1	0.6	0.73	1									
E18:0	0.8	0.72	0.7	1								
FAEEs	0.8	0.96	0.79	0.79	1							
ETG	0.4	0.59	0.56	0.45	0.62	1						
AST	0.1	0.08	-0.11	-0.08	0.06	-0.02	1					
ALT	0	0.04	-0.05	-0.1	0.04	-0.09	0.7	1				
GGT	0.2	0.19	0.07	0.06	0.2	0.1	0.46	0.51	1			
MCV	0.1	0.07	0.05	0.02	0.08	0.12	0.16	0.03	0.1	1		
CDT	0.3	0.24	0.26	0.24	0.29	0.31	-0.03	0	0.1	-0.05	1	
BMI	0.1	0.03	-0.01	0.07	0.05	0.04	0.18	0.32	0.4	-0.03	0.22	1

Table 3 Table representing the values related to FAEEs (as the sum of E14:0, E16:0, E18:0 and E18:1; [ng/mg]) and EtG [pg/mg] variables. The likelihood ratio values were calculated by means of the non-naïve LR model named as $LR_{FAEEs,EtG}$ (which takes into account FAEEs and EtG variables) and the LR model (LR_{LD}) obtained after the application of LDA and KDE procedures. Columns representing the final response of the LR models and the support that is delivered to the relative hypothesis, according to Table 4, are reported too.

Individuals	FAEEs (ng/mg)	EtG (pg/mg)	$LR_{FAEEs,ETG}$	Support to the hypothesis	Strength of the support	LR_{LD}	Support to the hypothesis	Strength of the support
case1	0.10	18	$2.0 \cdot 10^{3}$	H_1	S	$2.9 \cdot 10^{4}$	H_1	VS
case2	0.07	19	$5.2 \cdot 10^{3}$	H_1	S	$3.2 \cdot 10^{4}$	H_1	VS
case3	0.01	7	$7.5 \cdot 10^{9}$	H_1	VS	$2.6 \cdot 10^{8}$	H_1	VS
case4	0.01	18	$2.3 \cdot 10^{4}$	H_1	VS	$8.2 \cdot 10^{5}$	H_1	VS
case5	0.16	11	$6.4 \cdot 10^{7}$	H_1	VS	$1.3 \cdot 10^{6}$	H_1	VS
case6	0.30	15	$5.6 \cdot 10^{5}$	H_1	VS	$7.8 \cdot 10^{5}$	H_1	VS
case7	0.29	98	$1.4 \cdot 10^{-2}$	H_2	M	$6.9 \cdot 10^{-3}$	H_2	MS
case8	1.31	97	$1.0 \cdot 10^{-5}$	H_2	*VS*	$2.6 \cdot 10^{-4}$	H_2	S
case9	0.52	51	$4.2 \cdot 10^{-2}$	H_2	*M*	$3.2 \cdot 10^{-2}$	H_2	M
case10	0.60	31	$5.2 \cdot 10^{-4}$	H_2	*S*	$6.9 \cdot 10^{-3}$	H_2	MS
case11	1.01	29	$3.8 \cdot 10^{-4}$	H_2	*S*	$4.2 \cdot 10^{-4}$	H_2	S
case12	0.01	2769	$1.5 \cdot 10^{64}$	H_1	VS	$1.6 \cdot 10^{-4}$	H_2	S
case13	0.46	43	$8.8 \cdot 10^{-4}$	H_2	S	$8.4 \cdot 10^{-4}$	H_2	S
case14	1.25	12	$2.3 \cdot 10^{3}$	H_1	S	$4.0 \cdot 10^{3}$	H_1	S
case15	6.53	18	$2.5 \cdot 10^{-1}$	H_2	L	$3.7 \cdot 10^{-2}$	H_2	M
case16	0.92	21	$9.3 \cdot 10^{-1}$	H_1	M	$9.9 \cdot 10^{-1}$	H_2	M
case17	0.71	11	$1.8 \cdot 10^{2}$	H_1	MS	$3.1 \cdot 10^{2}$	H_1	MS
case18	11.98	26	$6.4 \cdot 10^{-3}$	H_2	MS	$1.4 \cdot 10^{-3}$	H_2	MS
case19	0.72	13	$5.1 \cdot 10^{2}$	H_1	MS	$1.7 \cdot 10^{2}$	H_1	MS
case20	0.96	13	$6.6 \cdot 10^{3}$	H_1	S	$9.0 \cdot 10^{3}$	H_1	S
case21	0.91	25	$1.0 \cdot 10^{-2}$	H_2	M	$2.0 \cdot 10^{-2}$	H_2	M
case22	0.57	27	$1.2 \cdot 10^{-2}$	H_2	M	$7.3 \cdot 10^{-2}$	H_2	M
case23	0.72	14	$4.0 \cdot 10^{2}$	H_1	MS	$5.6 \cdot 10^{2}$	H_1	MS
case24	2.75	22	$2.1 \cdot 10^{-2}$	H_2	M	$5.9 \cdot 10^{-2}$	H_2	M
case25	0.67	19	$9.8 \cdot 10^{3}$	H_1	S	$4.2 \cdot 10^{3}$	H_1	S
case26	1.51	14	$4.0 \cdot 10^{2}$	H_1	MS	$1.4 \cdot 10^{3}$	H_1	S
case27	0.65	13	$1.4 \cdot 10^{2}$	H_1	MS	$1.0 \cdot 10^{3}$	H_1	S
case28	0.54	6	$3.5 \cdot 10^{6}$	H_1	VS	$8.6 \cdot 10^{7}$	H_1	VS
case29	0.54	5	$1.3 \cdot 10^{8}$	H_1	VS	$1.1 \cdot 10^{5}$	H_1	VS
case30	0.63	7	$8.1 \cdot 10^{4}$	H_1	VS	$3.3 \cdot 10^{5}$	H_1	VS

case31	2.08	4	$6.6\cdot10^{7}$	H_1	VS	$5.6\cdot10^{6}$	H_1	VS
case32	0.58	3	$2.7\cdot10^{13}$	H_1	VS	$8.9\cdot10^{13}$	H_1	VS
case33	0.70	6	$4.4\cdot10^{5}$	H_1	VS	$1.5\cdot10^{7}$	H_1	VS
case34	0.62	9	$3.3\cdot10^{3}$	H_1	S	$3.0\cdot10^{4}$	H_1	VS
case35	2.08	4	$6.6\cdot10^{7}$	H_1	VS	$6.5\cdot10^{6}$	H_1	VS
case36	0.52	8	$6.3\cdot10^{4}$	H_1	VS	$5.1\cdot10^{4}$	H_1	VS
case37	5.15	6	$7.0\cdot10^{4}$	H_1	VS	$1.0\cdot10^{4}$	H_1	VS
case38	0.67	1.1	$3.1\cdot10^{32}$	H_1	VS	$1.3\cdot10^{27}$	H_1	VS
case39	0.52	2	$4.9\cdot10^{22}$	H_1	VS	$2.2\cdot10^{18}$	H_1	VS
case40	1.53	4	$7.6\cdot10^{7}$	H_1	VS	$6.2\cdot10^{7}$	H_1	VS
case41	6.99	1.4	$2.2\cdot10^{28}$	H_1	VS	$2.2\cdot10^{25}$	H_1	VS
case42	0.73	2	$3.2\cdot10^{21}$	H_1	VS	$1.6\cdot10^{21}$	H_1	VS
case43	1.03	6	$3.6\cdot10^{4}$	H_1	VS	$9.3\cdot10^{5}$	H_1	VS

Table 4 Table representing the verbal scale adopted, according to literature [28,43], in order to convert the LR values into the strength of support to be delivered to hypothesis indicated by the LR model to be the most probable.

Likelihood Ratios Ranges	**Verbal equivalent** • H_1: the subject is not a chronic alcohol abuser • H_2: the subject is a chronic alcohol abuser
$LR < 10^{-4}$	very strong (VS) support to H_2
$10^{-4} \leq LR < 10^{-3}$	strong (S) support to H_2
$10^{-3} \leq LR < 10^{-2}$	moderately strong (MS) support to H_2
$10^{-2} \leq LR < 10^{-1}$	moderate (M) support to H_2
$10^{-1} \leq LR < 1$	limited (L) support to H_2
$LR = 1$	inconclusive support to both hypotheses
$1 < LR \leq 10^{1}$	limited (L) support to H_1
$10^{1} < LR \leq 10^{2}$	moderate (M) support to H_1
$10^{2} < LR \leq 10^{3}$	moderately strong (MS) support to H_1
$10^{3} < LR \leq 10^{4}$	strong (S) support to H_1
$LR > 10^{4}$	very strong (VS) support to H_1