Data in Brief 12 (2017) 1–8

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Direct and indirect alcohol biomarkers data collected in hair samples - multivariate data analysis and likelihood ratio interpretation perspectives

Eugenio Alladio [a,b,*], Agnieszka Martyna [c], Alberto Salomone [b], Valentina Pirro [d], Marco Vincenti [a,b], Grzegorz Zadora [c,e]

[a] Dipartimento di Chimica, Università degli Studi di Torino, Via P. Giuria 7, 10125 Torino, Italy
[b] Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Regione Gonzole 10/1, 10043 Orbassano, Torino, Italy
[c] Department of Analytical Chemistry, Chemometric Research Group, Institute of Chemistry, The University of Silesia, Szkolna 9, 40-006 Katowice, Poland
[d] Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, 47907 Indiana, USA
[e] Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland

ARTICLE INFO

Article history:
Received 30 December 2016
Received in revised form
17 February 2017
Accepted 9 March 2017
Available online 16 March 2017

Keywords:
Alcohol
Likelihood ratio
Multivariate data analysis
Empirical cross entropy
Ethyl glucuronide
Fatty Acid Ethyl Esters
Hair analysis

ABSTRACT

The concentration values of direct and indirect biomarkers of ethanol consumption were detected in blood (indirect) or hair (direct) samples from a pool of 125 individuals classified as either chronic (i.e. positive) and non-chronic (i.e. negative) alcohol drinkers. These experimental values formed the dataset under examination (Table 1). Indirect biomarkers included: aspartate transferase (AST), alanine transferase (ALT), gamma-glutamyl transferase (GGT), mean corpuscular volume of the erythrocytes (MCV), carbohydrate-deficient-transferrin (CDT). The following direct biomarkers were also detected in hair: ethyl myristate (E14:0), ethyl palmitate (E16:0), ethyl stearate (E18:1), ethyl oleate (E18:0), the sum of their four concentrations (FAEEs, i.e. Fatty Acid Ethyl Esters) and ethyl glucuronide (EtG; pg/mg). Body mass index (BMI) was also collected as a potential influencing factor. Likelihood ratio (LR) approaches have been used to provide predictive

DOI of original article: http://dx.doi.org/10.1016/j.forsciint.2016.12.019
* Corresponding author at: Dipartimento di Chimica, Università degli Studi di Torino, Via Pietro Giuria 7, 10125 Torino, Italy.
E-mail addresses: ealladio@unito.it, eugenio.alladio@gmail.com (E. Alladio), ag.rzepecka@gmail.com (A. Martyna), alberto.salomone@antidoping.piemonte.it (A. Salomone), vpirro@purdue.edu (V. Pirro), marco.vincenti@unito.it (M. Vincenti), gzadora@ies.krakow.pl (G. Zadora).

http://dx.doi.org/10.1016/j.dib.2017.03.026

models for the diagnosis of alcohol abuse, based on different combinations of direct and indirect alcohol biomarkers, as described in "Evaluation of direct and indirect ethanol biomarkers using a likelihood ratio approach to identify chronic alcohol abusers for forensic purposes" (E. Alladio, A. Martyna, A. Salomone, V. Pirro, M. Vincenti, G. Zadora, 2017) [1].

Specifications Table

Subject area	*Chemistry*
More specific subject area	*Biomarkers of ethanol consumption in biological samples*
Type of data	*Tables, figures*
How data was acquired	*Analysis by Likelihood Ratio (LR) approach regarding the collected concentration values of the direct and indirect biomarkers of alcohol consumption.*
Data format	*Analyzed*
Experimental factors	*Correct classification rates and Empirical Cross Entropy (ECE) plots [2,3] were employed to evaluate LR models*
Experimental features	*AST, ALT and GGT were measured by means of colorimetric assays, MCV was measured with an on-purpose hematological auto-analyzer, %CDT was determined by an ad hoc High Performance Liquid Chromatography (HPLC) reagent kit, FAEEs were detected by HS-SPME-GC/MS analysis and EtG concentrations were monitored by Ultra High Performance Liquid Chromatography - Tandem Mass Spectrometry (UHPLC–MS/MS).*
Data source location	*Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Regione Gonzole 10/1, 10043 Orbassano, Torino, Italy.*
Data accessibility	*Data are included in this paper*

Value of the data

- The data reported here represent a valuable collection of all the common biomarkers of alcohol abuse used worldwide; the distinct populations of chronic and non-chronic alcohol consumers can possibly be used by other researcher to develop further interpretation models.
- The Empirical Cross Entropy plots provide a novel way to look at the effectiveness of alcohol biomarkers that other researcher may use for comparison with more traditional data representations.
- The detailed data report allows a clear comparison between univariate, multivariate and Bayesian approaches, where the latter is suggested as a benchmark for further developments.
- The mathematical background reported in the "materials and methods" section allows other researcher to transpose the offered approach to different applications.

1. Data

Data relative to the population of 125 individuals monitored, previously classified as either chronic (i.e. positive) and non-chronic (i.e. negative) alcohol drinker, are available in Table 1. Analysis of likelihood ratio models and its performance metrics, such as Empirical Cross Entropy plots (ECE), allowed to compare the predictive capabilities of direct and indirect biomarkers of ethanol consumption, as described in [1].