

Estudio de los factores asociados al uso de sustancias en la adolescencia mediante Reglas de Asociación

Study of the factors associated with substance use in adolescence using Association Rules

ELENA GERVILLA GARCÍA; BERTA CAJAL BLASCO;
RAFAEL JIMÉNEZ LÓPEZ; ALFONSO PALMER POL.

Universidad de las Islas Baleares

Enviar correspondencia a:

Alfonso Palmer. Universidad de las Islas Baleares
Ctra. Valldemossa Km. 7,5. 07122 Palma de Mallorca, Spain
alfonso.palmer@uib.es

recibido: Marzo 2010
aceptado: Septiembre 2010

RESUMEN

El objetivo del presente estudio es analizar los factores relacionados con el uso de sustancias adictivas en la adolescencia mediante reglas de asociación, herramientas descriptivas incluidas en Data Mining. Para ello se cuenta con una base de datos referidos al consumo de sustancias adictivas en la adolescencia y se utiliza el paquete *arules*, integrado en el programa de libre distribución R (versión 2.10.1). La muestra está formada por 9.300 estudiantes de edades comprendidas entre los 14 y los 18 años (47,1% chicos y 52,9% chicas) con una edad media de 15,6 años (SE=1,2). Los adolescentes contestaron un cuestionario anónimo que incluía preguntas sobre factores de riesgo personales, familiares y ambientales para el consumo de sustancias. Las mejores reglas obtenidas en relación al consumo de sustancias relacionan el consumo de alcohol con la educación paterna percibida y el consumo de los amigos (confianza= 0.8528), el consumo de tabaco, cannabis y cocaína con la actuación paterna percibida y la realización de conductas ilegales (confianzas de 0.8032, 0.8718 y 1.0000, respectivamente) y el uso de éxtasis con el consumo de los iguales (confianza = 1.0000). En general, las reglas de asociación muestran de forma sencilla la relación existente entre ciertas pautas de actuación paterna percibida, la emisión de conductas desviadas de las normas de comportamiento social, el consumo por parte del grupo de iguales y el abuso de drogas, legales e ilegales, en la adolescencia. Se describen las implicaciones de los resultados obtenidos así como la utilidad de esta nueva metodología de análisis.

Palabras clave: minería de datos, adolescencia, consumo de sustancias, reglas de asociación, abuso de drogas.

ABSTRACT

The aim of this study is to analyse the factors related to the use of addictive substances in adolescence using association rules, descriptive tools included in Data Mining. Thus, we have a database referring to the consumption of addictive substances in adolescence, and use the free distribution program in the R *arules* package (version 2.10.0). The sample was made up of 9,300 students between the ages of 14 and 18 (47.1% boys and 52.9% girls) with an average age of 15.6 (SE=1.2). The adolescents answered an anonymous questionnaire on personal, family and environmental risk factors related to substance use. The best rules obtained with regard to substance use relate the consumption of alcohol to perceived parenting style and peer consumption (confidence = 0.8528), the use of tobacco (smoking), cannabis and cocaine to perceived parental action and illegal behaviour (confidence = 0.8032, 0.8718 and 1.0000, respectively), and the use of ecstasy to peer consumption (confidence = 1.0000). In general, the association rules show in a simple manner the relationship between certain patterns of perceived parental action, behaviours that deviate from social behavioural norms, peer consumption and the use of different legal and illegal drugs of abuse in adolescence. The implications of the results obtained are described, together with the usefulness of this new methodology of analysis.

Key Words: data mining, adolescence, substance consumption, association rules, drugs of abuse.

INTRODUCTION

Rates of addictive substances use by Spanish adolescents are among the highest in Europe (European Monitoring Centre for Drugs and Drug Addiction, 2008). Some of the factors that have been related to the use of substances at this stage include certain personality traits, with deviant social behaviour standing out (Farrell, Sullivan, Esposito, Meyer and Valois, 2005; Graña, 2000; Jessor and Jessor, 1980). Moreover, family factors such as the way parents bring up their children (Beck, Boyle and Boekeloo, 2004; Becoña, 2002; Cava, Murgui and Musitu, 2008; De la Villa, Rodríguez and Sirvent, 2006; Espada and Méndez, 2002; Fernández, Secades, Vallejo and Errasti, 2003; Gerra, Angioni, Zaimovic, Moi, Bussandri, Bertacca et al., 2004; Martínez and Robles, 2001; Mewse, Eiser, Slater and Lea, 2004; Muñoz and Graña, 2001; Olsson, Coffey, Toumbourou, Bond, Thomas and Patton, 2003) or a "distant" relationship between parents and their children (Brook, Brook, De La Rosa, Whiteman, Jonson and Montoya, 2001) have also been shown to be related to substance use at this stage of development. Finally, the consumption of substances by the peer group has been repeatedly associated with the consumption of both legal and illegal substances in adolescence (Ciairano, Bosma, Miceli and Settani, 2008; Dick, Pagan, Viken, Purcell, Kaprio, Pulkkinen, et al., 2007; Kokkevi, Richardson, Florescu, Kuzman and Stergar, 2007; Scholte, Poelen, Willemsen, Boomsma and Engels, 2008; Simons-Morton, 2007).

In the field of prevention, it is crucial to identify the factors associated with the use of addictive substances, as well as the variables related to the absence of this consumption. As a result, many studies have analysed the data concerning substance use with the aim of discerning related psychosocial factors. Most research carries out this analysis using traditional statistical techniques, but there are also some studies which use statistical tools capable of discovering and expressing the relationships in the data in different ways. Within this framework, the interdisciplinary field of *Data Mining* arises from the confluence of areas such as statistics and artificial intelligence, and could be defined as the "discovery of interesting, unexpected or valuable structures in large databases" (Hand, 2007).

One of the most representative descriptive tools in *Data Mining* is Association Rules (ARs) (Agrawal, Imielinski and Swami, 1993). The aim of ARs is to identify patterns in data without knowing "a priori" about the reality. ARs are also referred to as "market basket analysis", since they originated with the study of databases of transactions customers made in order to find relationships between the purchase of different items. These "market baskets" consist of a bundle of operations carried out by customers, where each operation lists all the items bought by a person at one time. Thus, reveal which products usually go together in the basket. This type of information is extremely valuable for the strategic positioning of products in department stores or for planning special offers or promotions.

However, beyond the "market basket", the use of ARs has been generalised to any area in which great quantities

of stored information are available and it is necessary to explore regularities or associations between categorical elements. Moreover, despite the fact that ARs are not of a strictly predictive nature (due to the non-supervised learning basis), the very construction of the rules enables their use in the task of classifying data whose structure clearly presents a variable to be classified (Larose, 2005). It is this last perspective that forms the basis of the present study.

The aim of this study is to analyse the factors related to the use of addictive substances in adolescence using association rules. The advantages and disadvantages of the AR technique are analysed, conclusions are drawn and evaluations are made about the rules detected and their limitations in the substance use context in question.

METHOD

Participants

A random sample was drawn from schools (clusters) in the island of Mallorca, using 47 schools out of a total of 122. The sample is made up of 9,300 students aged 14 to 18 (47.1% boys and 52.9% girls). It is worth pointing out that this constitutes 41.16% of the population (N = 22,593) from which it was extracted. Average age of the adolescents is 15.6 years (SE = 1.2).

Procedure

The adolescents responded anonymously to a questionnaire for collecting information about frequency of use of different addictive substances and about certain psychosocial variables.

Currently, there are numerous programmes available which permit the extraction of ARs in a simple way. Nevertheless, of all these, the R program is that which stands out, as it is free and is continually being updated. The advantages of R have already been well described (Elosua, 2009), but we should also like to highlight its use within the methodology of *Data Mining*, and specifically for extracting ARs.

In this study the data are analysed using ARs (generated with the *arules* package of the R program, version 2.10.1).

Variables

Variables referring to how parents bring up their children, substance use by one's peer group and antisocial behaviour were taken into account, as well as gender.

The variables concerning the way parents bring up their children were collected by means of 30 questions on this subject, referring to: time parents spend with their teenagers; discipline or permissiveness; use of rewards and punishments; expression of affection; laying down of norms or rules; and

agreement between the two parents. Response options were "Never", "Sometimes" and "Always").

The information about antisocial behaviour is collected according to five dichotomous variables: "I cause commotions and kick up rows", "I do forbidden, illegal things", "I break, burn or damage other people's property", "I fight and insult others" and "I answer adults back".

Finally, substance use by the peer group is coded according to five categories. Thus, the adolescents are required to report how many of their friends consume each of the different substances analysed ("None", "Few", "Half", "Most" or "All").

Data Analysis

A fundamental distinction between the techniques included in *Data Mining* is that of supervised and non-supervised methods (Shmueli, Patel and Bruce, 2005). Supervised learning algorithms are used for classification and prediction where we have data available in which the result of interest is known, from which the algorithm "learns" about the relationship between the predictor variables and the result variable. On the other hand, non-supervised learning algorithms are those in which there is no result variable to predict or classify and the aim is to detect patterns or regularities in the data. Association rules are found in the latter perspective.

With the aim of finding interesting relationships between the data, ARs were calculated using the classical *Apriori* algorithm (Agrawal and Srikant, 1994). The basic idea of this algorithm is to progressively and recursively generate frequent itemsets, that is, sets of items that appear together in the database on a given (minimum) percentage of occasions.

ARs are made up of the antecedent (the first part of the rule) and the consequent (the second part), presenting the format "If A then C". The interestingness of the ARs was evaluated using the most well-known indexes that express their degree of uncertainty: support, confidence and lift (Han and Kamber, 2006).

Support is defined as the probability that an operation selected in a random way from the database contains all the items in the antecedent and the consequent, that is,

the number of transactions in the database that include all the items in the antecedent and consequent of the rule. Confidence is the conditional probability that an operation randomly selected from the database includes all the items in the consequent if this operation includes all the items in the antecedent. Finally, lift is the most popular interestingness measurement, and is expressed as the confidence rule divided by the confidence assuming the independence of the consequent and the antecedent. Lift indicates what probability we have of finding the consequent by limiting the search to those sets of items where the antecedent is present: the greater the lift, the greater the strength of association.

RESULTS

Below we present the results obtained using the *arules* package of the R program. In order to exemplify the results, only the three ARs showing the best lift indexes are shown in the tables.

It is worth noting that in the case of alcohol consumption we work with the whole matrix (N = 9,300); however, for the rest of the substances, due to the low prevalence of consumers (see Table 1), and with the aim of finding ARs that include both use and absence of use, samples were made up of users and non-users. Thus, for each of the substances, all the adolescents in the total sample who consumed the substance were selected and, afterwards, out of the total of non-consumers of the substance, a random sample of adolescents was taken to give a number equal to the sample of consumers. In Table 2 we can see the size of the samples analyzed with the prevalence of use of addictive substances in each sample.

Table 1. Prevalence of consumption of different substances in the sample (n = 9,300).

	Do not consume	Do consume	n
Alcohol	47.3%	52.7%	9300
Tobacco	75.0%	25.0%	4464
Cannabis	81.4%	18.6%	3302
Cocaine	98.4%	1.6%	278
Ecstasy	99.2%	0.8%	150

Table 2. Size of the samples analyzed in this study and prevalence of use of addictive substances in each sample.

		Alcohol (n = 9300)	Tobacco (n = 4464)	Cannabis (n = 3302)	Cocaine (n = 278)	Ecstasy (n = 150)
Alcohol	Do consume	52.7%	64.3%	67.0%	72.2%	68.9%
	Do not consume	47.3%	35.7%	33.0%	27.8%	31.1%
Tobacco	Do consume	25.0%	50.0%	43.3%	51.6%	49.3%
	Do not consume	75.0%	50.0%	56.7%	48.4%	50.7%
Cannabis	Do consume	18.6%	30.6%	50.0%	50.4%	47.3%
	Do not consume	81.4%	69.4%	50.0%	49.6%	52.7%
Cocaine	Do consume	1.6%	2.7%	3.6%	50.0%	41.3%
	Do not consume	98.4%	97.3%	96.4%	50.0%	58.7%
Ecstasy	Do consume	0.8%	1.5%	1.9%	22.7%	50.0%
	Do not consume	99.2%	98.5%	98.1%	77.3%	50.0%

The system generates 632,495 ARs that relate the variables studied in this work, 139 of which lead to alcohol use (that is, inform of alcohol use in the consequent of the AR) and 27 to the absence of alcohol use. The ARs obtained (Table 3) show that there is a relationship between alcohol consumption, paternal/maternal actions, consumption by the peer group and not damaging other people's property; whereas the absence of alcohol use is related to the absence of antisocial behaviour.

In the first AR concerning the use of alcohol, the interpretation of the rule is as follows: "In 10% of adolescents the three conditions are found together (the father sometimes disciplines, the mother never punishes for no reason and all the friends consume alcohol) and, moreover, 85% of these individuals also consume alcohol". A lift greater than 1 indicates that the AR found is useful.

Table 3. Association Rules with interestingness measures for alcohol consumption

Association rules	Support	Confidence	Lift
If the father sometimes disciplines and the mother never punishes for no reason and all the friends consume alcohol, then he/she consumes alcohol.	0.1003	0.8528	1.6617
If the father sometimes disciplines and never punishes for no reason and all the friends consume alcohol, then he/she consumes alcohol.	0.1055	0.8523	1.6607
If the father never disciplines and is sometimes permissive and all the friends consume alcohol, then he/she consumes alcohol.	0.1003	0.8505	1.6572
If few friends consume alcohol, he/she does not do forbidden, illegal things, does not break, burn or damage other people's property and does not fight or insult others, then he/she does not consume alcohol.	0.1104	0.8460	1.8365
If few friends consume alcohol, he/she does not do forbidden, illegal things, does not break, burn or damage other people's property and does not answer adults back, then he/she does not consume alcohol.	0.1039	0.8459	1.8363
If the mother never punishes for no reason, few friends consume alcohol, he/she does not do forbidden, illegal things and does not break, burn or damage other people's property, then he/she does not consume alcohol.	0.1030	0.8433	1.8307

As far as tobacco use (smoking) is concerned, the programme finds 532,497 ARs that relate the variables studied, 2 of which lead to smoking and 130 to non-smoking. The ARs (Table 4) show that smoking is related to some illegal behaviour whereas its absence is linked to certain patterns of parental action and to low levels of smoking by peers, as well as the absence of antisocial behaviours.

In the consumption of cannabis, the system generates 451,372 ARs, 107 of which lead to the consumption of cannabis and 266 to the absence of use. In the consumption of cocaine the programme finds 427,259 ARs, 248 of which lead to the consumption of cocaine and 7,043 to the absence of use. Cannabis and cocaine ARs (Tables 5 and 6) reveal a relationship between the use of these substances, perceived paternal/maternal action and deviant social behaviour.

Table 4. Association rules with interestingness measures for tobacco consumption

Association rules	Support	Confidence	Lift
If the mother never punishes for no reason and the adolescent does forbidden, illegal things, then he/she consumes tobacco.	0.1125	0.8032	1.6064
If he/she does forbidden, illegal things but does not break, burn or damage other people's property, then he/she consumes tobacco.	0.1201	0.8036	1.6072
If the mother never punishes for no reason, few friends smoke, he/she does not cause commotions or kick up rows, nor does forbidden, illegal things, then he/she does not consume tobacco.	0.1073	0.8615	1.7230
If the mother is always affectionate, few friends smoke, he/she does not do forbidden, illegal things, nor fight or insult others, then he/she does not consume tobacco.	0.1019	0.8601	1.7202
If the mother never punishes for no reason and the father does not impose senseless rules, few friends smoke and he/she does not do forbidden, illegal things, then he/she does not consume tobacco.	0.1026	0.8593	1.7176

Table 5. Association rules with interestingness measures for cannabis consumption

Association rules	Support	Confidence	Lift
If the mother sometimes disciplines and the adolescent does forbidden, illegal things, then he/she consumes cannabis.	0.1030	0.8718	1.7436
If the mother sometimes interferes in his/her business and the adolescent does forbidden, illegal things, then he/she consumes cannabis.	0.1033	0.8699	1.7398
If the mother never punishes for no reason, the father never imposes senseless rules and the adolescent does forbidden, illegal things, then he/she consumes cannabis.	0.1027	0.8670	1.7340
If the mother always tells him/her off when he/she does something wrong, no friend consumes cannabis, he/she does not do forbidden, illegal things and does not fight or insult others, then he/she does not consume cannabis.	0.1069	0.9832	1.9666
If the mother is always affectionate, no friend consumes cannabis, he/she does not do forbidden, illegal things and does not fight or insult others, then he/she does not consume cannabis.	0.1235	0.9831	1.9663
If the father never punishes for no reason, the mother is always affectionate, no friend consumes cannabis and he/she does not cause commotions or kick up rows, then he/she does not consume cannabis.	0.1029	0.9827	1.9653

Finally, for the use of ecstasy, the programme finds 400,576 ARs, 974 of which lead to the consumption of ecstasy and 13,628 to the absence of use. The ARs obtained (Table 7) reveal the association between the use of ecstasy, consumption by the peer group and the presence of deviant social behaviour, as well as paternal and maternal actions.

Table 6. Association rules with interestingness measures for cocaine consumption

Association rules	Support	Confidence	Lift
If the mother and father never punish for no reason, the adolescent is a girl and does forbidden, illegal things, then she consumes cocaine.	0.1151	1	2
If the father never punishes for no reason, the adolescent is a girl and does forbidden, illegal things, then she consumes cocaine.	0.1223	0.9714	1.9429
If the mother never imposes senseless rules, the father never interferes in his/her business and he/she does forbidden, illegal things, then he/she consumes cocaine.	0.1115	0.9687	1.9375
If the father is sometimes affectionate, no friend consumes cocaine and he/she does not fight or insult others, then he/she does not consume cocaine.	0.1043	1	2
If the father is sometimes affectionate, no friend consumes cocaine and he/she does not break, burn or damage other people's property, then he/she does not consume cocaine.	0.1115	1	2
If the father sometimes spends time with the adolescent and never shows a lack of agreement with the mother and no friend consumes cocaine, then he/she does not consume cocaine.	0.1043	1	2

Table 7. Association rules with interestingness measures for ecstasy consumption

Association rules	Support	Confidence	Lift
If all the friends consume ecstasy, then he/she consumes ecstasy.	0.1200	1	2
If all the friends consume ecstasy and he/she does forbidden, illegal things, then he/she consumes ecstasy.	0.1133	1	2
If the father and mother never value his/her opinion, then he/she consumes ecstasy.	0.1200	1	2
If the mother always spends time with the adolescent, the mother shows no lack of agreement with the father and no friend consumes ecstasy, then he/she does not consume ecstasy.	0.1000	1	2
If the mother sometimes imposes senseless rules, no friend consumes ecstasy and he/she does not do forbidden, illegal things, then he/she does not consume ecstasy.	0.1067	1	2
If the father does not show a lack of agreement with the mother, the father always values his/her opinion and no friend consumes ecstasy, then he/she does not consume ecstasy.	0.1200	1	2

Although the samples for the last two substances are smaller, it is worth noting that even small samples can give good association rules (see Kivinen and Mannila, 1994, for a general analysis of the relationship between the logical form of the discovered knowledge and the samples sizes needed for discovering it).

DISCUSSION

The aim of this study is to analyse the factors related to the use of addictive substances in adolescence using association rules. The ARs obtained show the relationship between how parents bring up their children (perceived by the adolescent), deviant social behaviour and the consumption of different legal and illegal substances. In addition, the relationship between the use of drugs in the peer group and their consumption by the adolescent can be observed. A perceived lack of parental discipline and antisocial behaviours are also related to the use of substances. On the other hand, perceived parental discipline associated with the expression of positive feelings and a group of friends who do not consume substances, together with the absence of antisocial behaviour, are the factors present in adolescents who say they do not consume substances.

The results obtained are in line with those of other national and international studies (Abar and Turrise, 2008; DiClemente, Wingood, Crosby, Sionean, Cobb, Harrington, et al., 2001; Espada, Pereira and García-Fernández, 2008; Martínez, Fuertes, Ramos and Hernández, 2003; Peña, Andreu and Graña, 2009; Stormshak and Dishion, 2009; Timmermans, Van Lier and Koot, 2008; Van der Vorst, Engels, Dekovic, Meeus and Versmulst, 2007; Walls, Fairlie and Wood, 2009).

Other studies have already related the aforementioned variables to the consumption of addictive substances; the novelty of the present study lies in using association rules to provide further insight into the factors involved in the use of these substances. This is a technique, indeed, which has been applied for some time in other areas of study. Put simply, ARs are the study of "what goes with what" and, in our work, they have enabled us to find out, for instance, what personality traits and what sort of upbringing may be related to the use of a certain substance.

ARs are simple, clear, useful, intuitive and intelligible tools that make it possible, in a descriptive way, to identify relationships between variables and, to date, studies using this tool in the Behavioural Sciences are few and far between. This technique permits the analysis of enormous numbers of variables, and each individual may have values in only a small number of the variables. Data can be highly heterogeneous. Nevertheless, the greatest disadvantage of ARs is probably the quantity of rules generated, so that considerable filtering is required by the user. Finally, it should be noted that finding an association rule does not necessarily imply that it is useful.

The great contribution of ARs as opposed to other classical statistical methods is the likelihood of discovering associations between items in a cumulative way, in the format "if-then". Thus, whereas linear regression or discriminant analysis focus on finding the weight of the predictor variables and require certain application conditions to be fulfilled, such as the normality of the variables (an assumption which is broken in many practical situations), ARs

make it possible to describe affinities between combinations of elements both in the antecedent and consequent of the rule. And all of this with objective numerical support that provides information concerning the quantity of transactions affected by the rule, their significance and the possibility that the rule has been produced by chance (in terms of support, confidence and lift, respectively). In this regard, if researchers wanted to obtain similar results using traditional data analysis techniques, they would have to carry out a simultaneous analysis of 68 contingency tables so as to be able to study the relationship between the set of variables analysed, which means a slow, arduous job, not exempt from the possibility of mistakes, when it comes to large dimension matrices. In our case there are 9.2×10^{63} combinations between the levels in the set of variables, which makes their analysis from a classical perspective practically impossible.

It is worth noting that ARs represent only "the tip of the iceberg" with respect to what *Data Mining* methodology can offer data analysis. A multitude of tools, both descriptive and predictive, are found in this methodology, providing new data analysis instruments which enable us to study the vast amounts of information we currently have from new perspectives (see Dunham, 2003; Han and Kamber, 2006; Shmueli et al., 2005). Some recent studies have used these tools in the field of addictive behaviours (Gervilla and Palmer, 2010; Gervilla, Cajal, Roca and Palmer, 2010), a field in which there is great potential for further research.

The present work provides evidence of the need for the design and implementation of preventive interventions for adolescents to target parental actions and deviant behaviour, as well as considering environmental risk factors.

This study had a large sample size which amounted to nearly 50% of the population studied, making it possible to investigate the use of a descriptive technique, included in *Data Mining*, capable of handling large volumes of data. Nevertheless, from a perspective related to substance use and with regard to the validity of the rules detected, it should be pointed out that the data analysed are self-informed, and do not take into account the views of different family members (Engels, Dekovic, Finkenauer and Meeus, 2001). In this regard, some authors have found that self-reports of drug use may be biased (Schwarz, 1999), so that adolescents report lower use than they would in an interview situation (Stone and Latimer, 2005).

ACKNOWLEDGEMENTS

This work was carried out thanks partly to support from the National Plan on Drugs (INT/2012/2002), as part of a three-year research project. Permission was obtained for all the research from the institutions and teachers and from the students themselves, whose participation was voluntary.

REFERENCES

- Abar, C. & Turrisi, R. (2008). How important are parents during the college years? A longitudinal perspective of indirect influences parents yield on their college teens' alcohol use. *Addictive Behaviors*, 33, 1360-1368.
- Agrawal, R. & Srikant, R. (1994). *Fast algorithms for mining association rules*. Paper presented at the 20th International Conference on Very Large Databases, Santiago, Chile.
- Agrawal, R., Imielinski, T. & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM-SIGMOD International Conference on Management of Data*, 207-216.
- Beck, K.H., Boyle, J.R. & Boekeloo, B.O. (2004). Parental monitoring and adolescent drinking: results of a 12-year follow-up. *American Journal of Health Behavior*, 28, 272-279.
- Becoña, E. (2002). Factores de riesgo y protección familiar para el uso de drogas. In R. Secades and J.R. Fernández (Eds.), *Intervención Familiar en la Prevención de las Drogodependencias*. Madrid: Plan Nacional sobre Drogas.
- Brook, J.S., Brook, D.W., De La Rosa, M., Whiteman, M., Jonson, E. & Montoya, I. (2001). Adolescent illegal drug use: the impact of personality, family, and environmental factors. *Journal of Behavioral Medicine*, 24, 183-203.
- Cava, M.J., Murgui, S. & Musitu, G. (2008). Diferencias en factores de protección del consumo de sustancias en la adolescencia temprana y media. *Psicothema*, 20, 389-395.
- Ciairano, S., Bosma, H.A., Miceli, R. & Settani, M. (2008). Adolescent substance use in two European countries: Relationships with psychosocial adjustment, peers, and activities. *International Journal of Clinical and Health Psychology*, 8, 119-138.
- De la Villa, M., Rodríguez, F.J. & Sirvent, C. (2006). Factores relacionados con las actitudes juveniles hacia el consumo de alcohol y otras sustancias psicoactivas. *Psicothema*, 18, 52-58.
- Dick, D.M., Pagan, J.L., Viken, R., Purcell, S., Kaprio, J., Pulkkinen, L., & Rose, R.J. (2007). Changing environmental influences on substance use across development. *Twin Research and Human Genetics*, 10, 315-326.
- DiClemente, R.J., Wingood, G.M., Crosby, R., Sionean, C., Cobb, B.K., Harrington, K.,... Oh, M.K. (2001). Parental monitoring: association with adolescents' risk behaviors. *Pediatrics*, 107, 1363-1368.
- Dunham, M.H. (2003). *Data Mining. Introductory and Advanced Topics*. New Jersey: Pearson Education.
- Elosua, P. (2009). ¿Existe vida más allá del SPSS? Descubre R. *Psicothema*, 21, 652-655.
- Engels, R.C.M.E., Dekovic, M., Finkenauer, C. & Meeus, W. (2001). Parental attachment and adolescents' and emotional adjustment: the associations with the social skills and relational competence. *Journal of Counseling Psychology*, 48, 428-439.
- Espada, J.P. & Méndez, F.X. (2002). Factores familiares, comportamientos perturbadores y drogas en la adolescencia. In J.R. Fernández and R. Secades (Eds.), *Intervención Familiar en la Prevención de las Drogodependencias*. Madrid: Plan Nacional sobre Drogas.
- Espada, J.P., Pereira, J.R. & García-Fernández, J. (2008). Influencia de los modelos sociales en el consumo de alcohol de los adolescentes. *Psicothema*, 20, 531-537.
- European Monitoring Centre for Drugs and Drug Addiction (2008). *2008 Annual Report: The state of the drugs problem in Europe*.

- [Document WWW]. URL Retrieved 8, November, 2008 from http://www.emcdda.europa.eu/attachements.cfm/att_64227_EN_EMCDDA_AR08_en.pdf
- Farrell, A. D., Sullivan, T. N., Esposito, L. E., Meyer, A. L. & Valois, R. F. (2005). A latent growth curve analysis of the structure of aggression, drug use, and delinquent behaviors and their interrelations over time in urban and rural adolescents. *Journal of Research on Adolescence, 15*, 179-204.
- Fernández, J.R., Secades, R., Vallejo, G. & Errasti, J.M. (2003). Evaluation of what parents know about their children's drug use and how they perceive the most common family risk factors. *Journal of Drug Education, 33*, 334-350.
- Gerra, G., Angioni, L., Zaimovic, A., Moi, G., Bussandri, M., Bertacca, S.,... Nicoli, M.A. (2004). Substance use among high-school students: relationships with temperament, personality traits, and parental care perception. *Substance Use & Misuse, 39*, 345-367.
- Gervilla, E. & Palmer, A. (2010). Prediction of cannabis and cocaine use in adolescence using decision trees and logistic regression. *The European Journal of Psychology Applied to Legal Context, 2*, 19-35.
- Gervilla, E., Cajal, B., Roca, J. & Palmer, A. (2010). Modelling alcohol consumption during adolescence using Zero Inflated Negative Binomial and Decision Trees. *The European Journal of Psychology Applied to Legal Context, 2*, 145-159.
- Graña, J.L. (2000). Factores psicológicos de riesgo y de protección para el consumo de drogas en adolescentes. *Psicología Conductual. Revista Internacional de Psicología Clínica de la Salud, 8*, 249-269.
- Han, J. & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd. ed.). San Francisco: Morgan Kaufmann.
- Hand, D.J. (2007). Principles of Data Mining. *Drug Safety, 30*, 621-622.
- Jessor, R. & Jessor, S. (1980). A social-psychological framework for studying drug use. In D.J. Letteri, M. Sayers and H.W. Pearson, eds. *Theories on Drug Abuse: Selected Contemporary Perspectives*, pp. 95-101. Rockville, MD: National Institute on Drug Abuse.
- Kivinen, J. & Mannila, H. (1994). The power of sampling in Knowledge Discovery. In *Proceedings of the Thirteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS'94)* (pp. 77-85). Minneapolis, MN: ACM Press.
- Kokkevi, A., Richardson, C., Florescu, S., Kuzman, M. & Stergar, E. (2007). Psychosocial correlates of substance use in adolescence: A cross-national study in six European countries. *Drug and Alcohol Dependence, 86*, 67-74.
- Larose, D.T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, NJ: Wiley.
- Martínez, J.L., Fuertes, A., Ramos, M. & Hernández, A. (2003). Consumo de drogas en la adolescencia: importancia del afecto y la supervisión parental. *Psicothema, 15*, 161-166.
- Martínez, J.M. & Robles, L. (2001). Variables de protección ante el consumo de alcohol y tabaco en adolescentes. *Psicothema, 13*, 222-228.
- Mewse, A.J., Eiser, J.R., Slater, A.M. & Lea, S.E.G. (2004). The smoking behaviors of adolescents and their friends: do parents matter? *Parenting: Science and Practice, 4*, 51-72.
- Muñoz, M. & Graña, J.L. (2001). Factores familiares de riesgo y de protección para el consumo de drogas en adolescentes. *Psicothema, 13*, 87-94.
- Olsson, C., Coffey, C., Toumbourou, J., Bond, L., Thomas, L. & Patton, G. (2003). Family risk factors for cannabis use: a population-based survey of Australian secondary school students. *Drug and Alcohol Review, 2*, 143-152.
- Peña, M. E., Andreu, J. M. & Graña, J. L. (2009). Multivariate model of antisocial behavior and substance use in Spanish adolescents. *Journal of Child & Adolescent Substance Abuse, 18*, 207-220.
- Scholte, R.H.J., Poelen, E.A.P., Willemsen, G., Boomsma D.I. & Engels, R.C.M.E. (2008). Relative risks of adolescent and young adult alcohol use: The role of drinking fathers, mothers, siblings, and friends. *Addictive Behaviors, 33*, 1-14.
- Schwarz, N. (1999). Self-report: how the questions shape the answers. *American Psychologist, 54*, 93-105.
- Shmueli, G., Patel, N.R. & Bruce, P.C. (2007). *Data mining in Excel: Lecture Notes and Cases*. Arlington, VA: Resampling Stats, Inc.
- Simons-Morton, B. (2007). Social influences on adolescent substance use. *American Journal of Health Behavior, 31*, 672-684.
- Stone, A.L. & Latimer, W.W. (2005). Adolescent substance use assessment: concordance between tools using self-administered and interview formats. *Substance Use & Misuse, 40*, 1865-1874.
- Stormshak, E.A. & Dishion, T.J. (2009). A school-based, family-centered intervention to prevent substance use: the family check-up. *The American Journal of Drug and Alcohol Abuse, 35*, 227-232.
- Timmermans, M. Van Lier, P.A.C. & Koot, H.M. (2008). Which forms of child/adolescent externalizing behaviors accounts for late adolescent risky sexual behaviour and substance use? *Journal of Child Psychology and Psychiatry, 49*, 386-394.
- Van der Vorst, H., Engels, R.C.M.E., Dekovic, M., Meeus, W. & Versmulst, A.A. (2007). Alcohol-specific rules, personality and adolescents' alcohol use: a longitudinal person-environment study. *Addiction, 102*, 1064-1075.
- Walls, T.A., Fairlie, A.M. & Wood, M.D. (2009). Parents do matter: a longitudinal two-part mixed model of early college alcohol participation and intensity. *Journal of Studies on Alcohol and Drugs, 70*, 908-918.

