

Classification Trees in a Casework Concerning Drug Addiction and Psychiatric Pathology

Uso degli Alberi di Classificazione in uno Studio su Tossicodipendenza e Malattia Psichiatrica

Patrizia Giaquinto

ISTAT,

Servizio MTS

Via Cesare Balbo 16, 00184 Roma

e-mail: patrizia_giaquinto@yahoo.it

Alessio Pollice

Dipartimento di Scienze Statistiche, Università degli Studi di Bari

Via C. Rosalba 53, 70124 Bari

e-mail: apollice@dss.uniba.it

Alessandra Dipietro

Dipartimento di Scienze Neurologiche e Psichiatriche,

Università degli Studi di Bari, P.zza G. Cesare 11, 70124 Bari

e-mail: panicolab@psichiat.uniba.it

Abstract: L'utilizzo degli alberi decisionali è ormai pratica consolidata nella ricerca epidemiologica al fine di identificare gruppi omogenei di pazienti (Koss e Feinstein, 1971), servendosi di modelli grafici. Nel presente lavoro tale metodologia viene applicata ad un insieme di dati clinici, rilevati in occasione di un progetto di ricerca riguardante la doppia diagnosi in psichiatria, intendendo quest'ultima come la presenza contemporanea di tossicodipendenze e gravi patologie mentali (De Leon, 1989). La finalità del lavoro è quella di classificare gli individui affetti da doppia diagnosi in base ad alcuni predittori ritenuti rilevanti.

Keywords: Psychiatric Epidemiology, Drug Addiction, CART.

1. Introduction: dual disorders and the Double Spiral project

The expression "dual disorder" was coined in the late 80's and means the concurrent presence of a serious mental illness with drug abuse or addiction (De Leon, 1989; Solomon, 1996). The importance of such disorders is agreed to be increasing in the last few years, with estimates of the prevalence of users of psychotropic substances ranging between 15 and 18% in the general population and between 19 and 30% for psychiatric patients (Galanter and Kleber, 1998). The development and testing of a therapeutic-rehabilitative model for patients suffering from dual disorders was furthered within the EQUAL EU program by a partnership among the two Departments of Neurology and Psychiatry and of Education and Didactics of the University of Bari with a local sanitary authority (ASL Ba3) and two private institutions in the field of voluntary services. The main objective of the Double Spiral initiative (IT-G-PUG-0111) is to adjust professional skills and organizational systems connected with the care and rehabilitation of substance use disorders and psychiatric disorders to the needs of patients affected by both pathologies. This paper is concerned with the first step of the project, aimed at ascertaining the size and the main features of the dual disorder phenomenon in the district

of Bari, in order to characterize a profile of necessary and appropriate abilities for its management.

2. Data Collection and Database Construction

The primary objective of the epidemiological study was thus to quantify the prevalence of psychiatric disorders belonging to axes I and II of DSM-IV-TR in patients with substance use disorders entering daytime social services (SerT) within the five local sanitary authorities of the Bari district (ASL Ba1-5) and living in private voluntary based therapeutic communities thereby. Psychiatric diagnoses were performed by the Mini International Neuropsychiatric Interview Plus and the Self-Report Symptom Inventory (SCL 90). As a secondary objective the population features referred to social, demographic and clinical aspects were measured by a specially made questionnaire. The nature and severity of problems caused by the substance use was also evaluated by a 142 items test leading to an Addiction Severity Index (ASI) referred to various relevant functional aspects of the life of patients. Finally, the level of disablement due to psychiatric disorders in the social, working, and family life was measured by a test on a standard disability scale (DISS). Due to its observational/surveillance nature the study was conducted according to a cross-sectional design, with the data collection taking place between 10 october 2002 and 10 january 2003. All patients of age attending the five social services and therapeutic communities in the reference period were asked to answer to a structured interview including a questionnaire and various tests. Non-compliance caused the exclusion from the study together with the presence of serious medical diseases and the treatment with drugs known to induce psychiatric pathologies. Interviewers were identified as medical doctors, psychiatrists and psychologists, belonging to the social services and to the Department of Neurology and Psychiatry of the University of Bari, who were expressly trained for the purpose. At the end of the data collection 691 eligible patients were interviewed, 256 of which were affected by dual disorders.

3. Classification Trees in Epidemiology

Decision trees are one of the most powerful techniques engaged in statistics for classification purposes and are successfully used in Epidemiology because of their inductive nature, which resembles clinicians' way of thinking in grouping similar behaviours or symptoms for observed individuals. Some examples of the method concern identifying high risk for myocardial infarction, evaluating chest pain, dental caries, diabetes, asthma, orthopaedic fractures and others (Marshall, 2001).

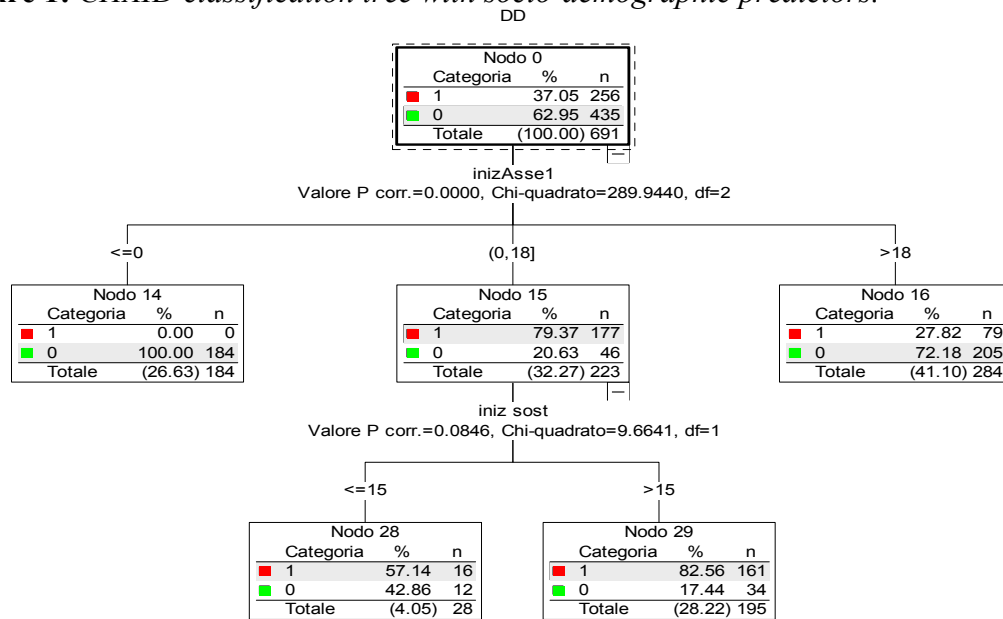
A decision tree represents a process of obtaining general models from real situations in order to learn how to predict (classify) objects in reference to an outcome variable, starting by an analysis of instances subdivided in known groups (predictions or classes). Instances usually consist of attributes describing the individuals and are supposed to be associated with the outcome variable and trees are generated by repeatedly splitting a sample of units according to their values. The final result is a graphical model of the mappings created for all the significant relationships between instances and classes (predictions). For details on the method see, for instance, Breiman *et al.*, 1984.

In our study the outcome variable is represented by the dual disorder diagnosis, assuming only two values, where 1 indicates the presence of dual disorder and 0 its absence, while the total number of possible instances is 190. The purpose of applying classification trees to the complete data set of 691 patients is therefore to exhibit the most significant variables concurrent to discriminate among those affected by dual disorders and those who are not. The next section shows some results obtained by classifying the entire sample using different subgroups of predictors and algorithms.

4. Some Numerical Results

We performed our experiment using AnswerTree 3.1 in SPSS package on a MS Windows XP platform. As our outcome variable is nominal, we could apply each of the four algorithms implemented: CHAID, exhaustive CHAID, CART and QUEST. We did several trials setting the stopping criterion for the expansion of the tree as follows: maximum depth fixed at 5 levels, minimum number of cases respectively equal to 25 for each 'parent' node and 5 for each 'child' node. At first, we used different combinations of predictors as it made sense to point out not only the most important variables in the study but also their interactions. After many comparisons similar results were obtained applying mainly the exhaustive CHAID and CART methods, in all cases the most considerable predictors being some mental symptoms (DDAI and GAD) related to the age of onset for axis I disorders and the use of stimulants. In order to emphasise the different aspects of the phenomenon we then decided to retain three groups of instances relevant to Psychiatric diagnoses, to socio-demographical variables and to psychotropic substances. First of all, by choosing Psychiatric diagnoses as predictors we obtained that out of 31 variables, indicating as many kinds of humour, anxiety, somatization and other disorders, only four of them were determinant at classification purposes. The first of these variables (DDAI) is the one which represents attention deficit or hyperactive persons, and mostly discriminates for the outcome variable. In fact, 78% of the cases not affected by DDAI in our study does not belong to the dual disorder group, while, on the contrary, only the remaining 22% has a dual disorder. The reason for this result can be explained by the fact that such a disorder is able to mask other serious mental troubles, so that an erroneous diagnosis in this sense, i.e., its absence, causes the exclusion of a dual disorder diagnosis too. The second relevant variable in this application is the social aversion followed by the generalized anxiety disorder (GAD) and the obsessive-compulsive disorder (DOC), which are all kind of neurosis, typical of drug addiction. Following the path of the tree, 85% out of the subjects who have neither social aversion nor attention deficit do not belong to the dual disorder class; for the GAD disorder the situation is similar to the previous variable, in fact, 88% of individuals without GAD disorder and not affected by social aversion, are not included in the case of dual disorder, and, finally, the most discriminating variable turns out to be the DOC one, because 90% of cases with none of the four diagnoses do not have dual disorders. The second group of variables considered in the study was that of 14 socio-demographical attributes, such as working conditions, income, education and age of disorder along axis I, ages of troubles related to substances and age of first abuse. In the first level of the tree (see Figure 1) we noted an important subdivision between those who exhibit axis I disorders before the age of 18 and the others (the zero age class refers to patients not ascribing to axis I).

Figure 1: CHAID classification tree with socio-demographic predictors.



The results suggest an important probability (79%) of being included in the dual disorder class for subjects under age, in comparison to over 18, the probability of the last ones going down to 28%. Furthermore, there is another distinction about under age, according to their age of first addiction: if they start abusing toxic substances over 15 years age, they have a much higher probability (83%) of belonging to the dual disorder patients. The last group of variables in the study took into account the different substances of abuse or addiction, surveyed by the MINI test. The most significant variables are about lifetime dependencies and three main substances. The use of stimulants seems to distinguish almost well for the outcome variable, as 64% of those who do not assume stimulant substances are not in the dual disorder class. Moreover, 68% of individuals using heroin and stimulants at the same time are affected by dual disorders. 78% of those who are not heroin addicts and rather use stimulants belong to the dual disorder class in any case. There also exists a certain probability (63%) of subjects who are in the dual disorder group and do not use stimulants but use tranquillizers in substitution.

References

- Breiman L, Freidman JH, Olshen RA, Stone CJ (1984) *Classification and regression trees*, Belmont, CA: Wadsworth International.
- De Leon G. (1989) Psychopathology and substance abuse and psychiatric disorders: what is being learned from research in therapeutic community, *Journal of Psychoactive Drugs*, 21, 177-188.
- Galanter M., Kleber H.D. (1998) *Trattamento di disturbi da uso di sostanze*, Masson, Milano.
- Koss N., Feinstein AR. (1971) *Computer aided prognosis. II. Development of a prognostic algorithm*. Arch Int Med, 127, 448-59.
- Marshall, RJ. (2001) The use of classification and regression trees in clinical epidemiology, *J. of Clinical Epidemiology*, 54, 603-609.
- Solomon J. (1996) Doppia diagnosi, *Personalità/Dipendenze*, 2, 279-289.