Latent Variable Mixture Modelling of Treated Drug Misuse in Ireland

Paul Cahill and Brendan Bunting[1]

Abstract

This study provides analyses and profiles of illegal drug usage in the Republic of Ireland. Two questions are addressed: a) can individuals be grouped into homogeneous classes based upon their type of drug consumption, and b) how do these classes differ in terms of other key background variables? The data reported in this study is from the National Drug Treatment Reporting System database in the Republic of Ireland. All analyses were carried out in collaboration with the Drug Misuse Research Division (the Irish REITOX / EMCDDA focal point). This database contains information on all 6994 individuals who received treatment for drug problems in the Republic of Ireland during 2000. The analysis was conducted in four steps. First, a single class model was examined in order to establish the respective probability associated with each drug type. Second, a series of unconditional latent class models was examined. This was done to establish the optimal number of latent classes required to describe the data, and to establish the relative size of each latent class. From this analysis the conditional probabilities for each individual, within a given class, were examined for typical profiles. Third, a series of conditional models was then examined in terms of key predictors (age and early school leavers). This analysis was conducted using MPlus 2.13. In the final stage of the research, the parameter estimates obtained from the multinomial logistic regression model (that was previously used to express the probability of an individual being in a given latent class, conditional on a series of covariates) were graphically modelled within EXCEL and the respective functions described. The results from this analysis will be described in terms of a) the profiling of typical serious drug misuse in Ireland, b) the clustering of drug types and, c) the respective importance of key background variables. The various profiles obtained are discussed in terms of health care strategies in Ireland.

---

[1] School of Psychology, University of Ulster at Magee Campus, Northern Ireland; p.cahill@ulster.ac.uk; bp.bunting@ulster.ac.uk

1   Introduction

The constant battle between the competing needs of health-care and social resources necessary for effective drug treatment has been apparent for the last three decades. These resources are not only financial: people, facilities, expertise and even time are all finite, and the most efficient use of them needs to be found within the drug-treatment services.

The underlying constructs of drug misuse and its characteristics can only be investigated indirectly using observable measures and indicators. Latent variable mixture modelling allows for unobserved heterogeneity in the sample (Muthén and Muthén, 2003) where different individuals can be seen to belong to different subpopulations, and can uncover some of the unique characteristics of this hidden and vulnerable population.

Over the last three decades large scale evaluation studies Drug Abuse Reporting Program (DARP) (Simpson and Sells, 1990), and the National Treatment Outcome Research Study (NTORS) (Gossop, Martin, and Stewart, 2003) have examined the effectiveness of drug treatment, in both the UK and US. Despite these considerable efforts, addiction-treatment researchers are still confronted with many of the same core questions as those of thirty years ago. These questions include:

- can individuals be grouped into homogeneous classes based upon their type of drug consumption, and
- how do these classes differ in terms of other key background variables?

2   Method

2.1   Participants

Participants were 6994 individuals who received treatment for problem drug use in the Republic of Ireland during the 2000 calendar year. Data returns for the 6994 clients attending treatment services during 2000 were provided by 154 treatment services (O'Brien, Kelleher, and Cahill, 2002).

2.2   Measures

The National Drug Treatment Reporting System (NDTRS) is an epidemiological database on treated drug misuse in the Republic of Ireland. The reporting system was originally developed in line with the Pompidou Group's Definitive Protocol

and subsequently refined in accordance with the Treatment Demand Indicator Protocol (European Monitoring Centre for Drugs and Drug Addiction & Pompidou Group, 2000).

Drug treatment data are viewed as an indirect indicator of drug misuse. These data are used at national and European levels to provide information on the characteristics of clients entering treatment, and on patterns of drug misuse, such as types of drugs used and consumption behaviours. They are "*valuable from a public health perspective to assess needs... and to plan and evaluate services*" (EMCDDA, 2002: 23).

The monitoring role of the NDTRS is recognised by the Irish Government, and data collection for the National Drug Treatment Reporting System is one of the actions identified by the government for implementation by all health authorities: "*All treatment providers should co-operate in returning information on problem drug use to the NDTRS*" (Department of Tourism, Sport and Recreation 2001: 118).

2.3   Analyses

The basic premise of the latent class analyses, relevant to this study is that the covariation found among the observed variables is due to each observed variable's relationship to the latent variable (Hagenaars and McCutcheon, 2002).
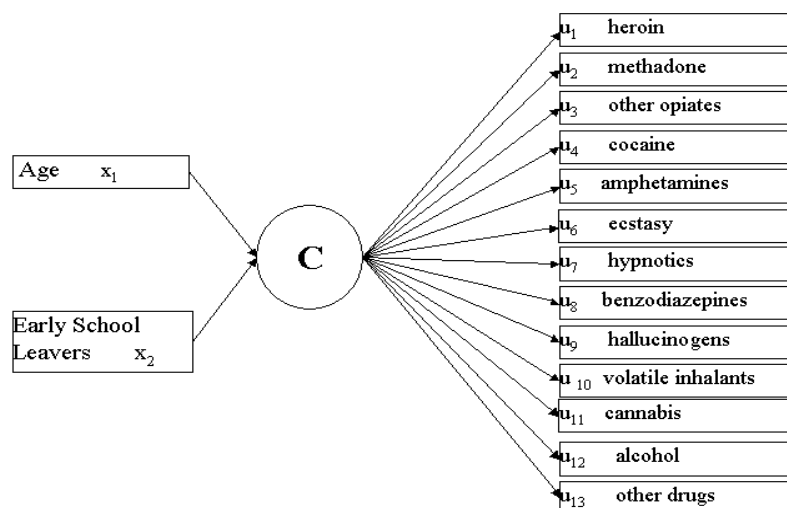

Figure 1:   Latent classes with covariates.

In this study, the observed drug consumption behaviours (u) are dichotomous in nature, representing the presence or absence of the particular drug in each

individual's drug using career. 'C' denotes the latent classes to be determined. Once the optimal number of latent classes is calculated, the key background variables of 'age' (continuous) and 'early school leavers' (dichotomous) are introduced.

The Bayesian Information Criterion (BIC) together with the Akaike Information Criterion (AIC) were used to determine the optimal number of unconditional latent classes necessary to adequately fit the data. The Bayesian information criterion, as proposed by Schwarz (1978), calculates the criterion for one or more fitted model objects for which a log-likelihood value can be obtained.

$$-2*log\text{-}likelihood + npar*log(nobs) \tag{1.1}$$

where *npar* represents the number of parameters and *nobs* the number of observations in the fitted model. BIC is a "parsimony criteria" used to perform model comparison and discrimination.

A series of conditional models was examined in terms of key predictors. Multinomial logistic regression is an extension of binary logistic regression that allows for the comparison of more than one contrast, simultaneously estimating the log odds of a number of contrasts. Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable (the natural log of the odds of the dependent occurring or not). In this way, logistic regression estimates the probability of a certain event occurring.   In the multinomial logit model it is assumed that the log-odds of each response follow a linear model. The logit model using the baseline-category logits with predictors **x** has the form

$$\log \frac{\pi_j(\mathbf{x})}{\pi_J(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}'_j \mathbf{x}, \tag{2.1}$$

where α_j is a constant and $\boldsymbol{\beta}'_j$ is a vector of regression coefficients of **x**, for J = 6, j = 1, 2, …, J-1, (Agresti, 2002).

This model is analogous to a logistic regression model, except that the probability distribution of the response is multinomial instead of binomial and we have J-1 equations instead of one. The multinomial logit model may also be written in terms of the original probabilities π_{ij} rather than the log-odds. Starting from Equation 2.1, then

$$\pi_j(\mathbf{x}) = \frac{\exp\left(\alpha_j + \boldsymbol{\beta}'_j \mathbf{x}\right)}{1 + \sum_{k=1}^{J-1} \exp\left(\alpha_k + \boldsymbol{\beta}'_k \mathbf{x}\right)} \tag{2.2}$$

where **x** are the explanatory variables age (years) and early school leavers, and j = 1, 2,… 5 for each of the six latent classes respectively. The denominator

remains the same for each probability, and the numerators for all *j* sum to the denominator, $\sum_j \pi_j(\mathbf{x}) = 1$. The parameters for the baseline category in the logit expressions equal to zero, therefore the model with category J as the baseline for the demoninator (equation 2.1), has $\alpha_J = \boldsymbol{\beta}_J = \mathbf{0}$, (Hosmer and Lemeshow, 2000).

2.4   Instruments / Psychometric software

Database preparation and data cleaning were completed using SPSS (11.0). There were no missing data. MPlus 2.13 was used, and all analyses were carried out in a single-step. MSExcel 2000 was used graphically to illustrate the multinomial logistic regression.
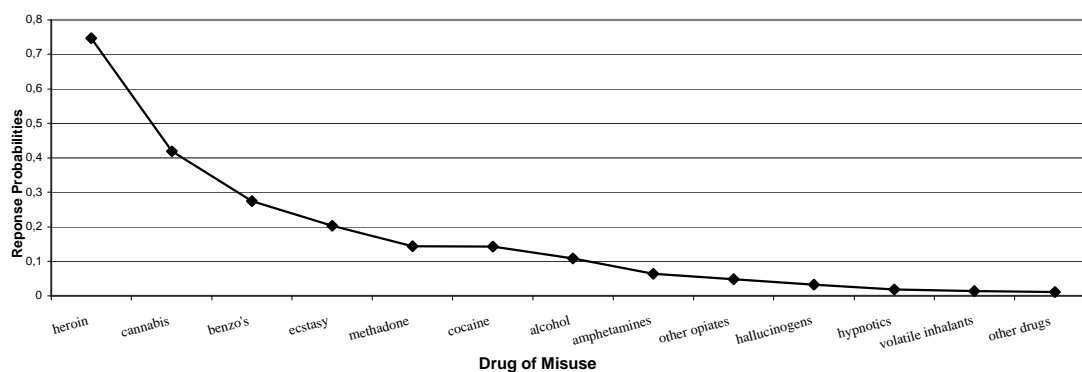
3   Results

Figure 2:   Positive response probabilities for the thirteen drug types.

Figure 2 illustrates the respective probabilities associated with each of the ten drug types among individuals receiving drug treatment during 2000. The majority of those seeking treatment (75%) reported heroin as one of their drugs of misuse. Cannabis was the next most probable drug of misuse among those seeking treatment (42%), and only one percent of those receiving treatment in 2000 reported use of volatile inhalants.
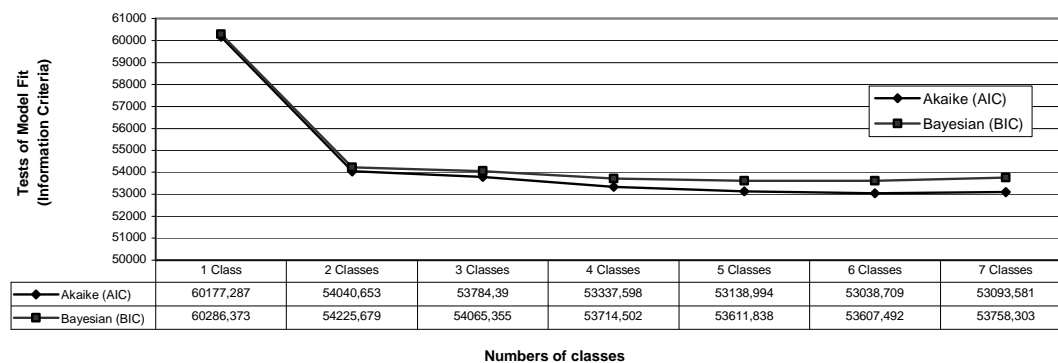
Figure 3: Inf. Criterion (BIC&AIC) testing model fit for seven model-solutions.

Table 1: Six latent classes, the average class probabilities by class and the associated probabilities of problem drug use.

		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**
Class	n =	950	577	1013	999	183	3272
size	% =	13.6	8.3	14.5	14.3	2.6	46.8

Average Class Probabilities by Class

	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**
Class 1	0.759	0.213	0.026	0.001	0.001	0.000
Class 2	0.099	0.847	0.026	0.000	0.023	0.005
Class 3	0.028	0.023	0.784	0.000	0.029	0.136
Class 4	0.001	0.000	0.010	0.989	0.000	0.000
Class 5	0.004	0.053	0.069	0.000	0.842	0.032
Class 6	0.001	0.005	0.141	0.000	0.035	0.818

Probabilities of Problem Use of Drug

Misuse of:

	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**
Heroin	0.065	0.079	0.973	0.856	0.517	1.000
Methadone	0.010	0.000	0.008	1.000	0.000	0.000
Other Opiates	0.000	0.015	0.053	0.038	0.397	0.033
Cocaine	0.215	0.028	0.292	0.118	0.068	0.097
Amphetamines	0.429	0.021	0.060	0.009	0.000	0.000
Ecstasy	0.842	0.363	0.319	0.047	0.012	0.000
Hypnotics	0.006	0.110	0.017	0.015	0.026	0.001
Benzo's	0.018	0.033	0.370	0.356	0.665	0.296
Hallucinogens	0.168	0.018	0.049	0.006	0.000	0.000
Volatile inhalants	0.007	0.101	0.008	0.003	0.000	0.000
Cannabis	0.883	0.930	0.654	0.250	0.216	0.122
Alcohol	0.391	0.365	0.014	0.022	0.248	0.018
Other drugs	0.010	0.008	0.006	0.011	0.082	0.007

	Clubbers	*Novices*	*Polydrug users*	*Methadone misusers*	*Benzo misusers*	*Heroin misusers*

3.1   Goodness of fit

The AIC and BIC fit indices improve as additional latent classes are added to the model, (for example, a three-class solution BIC value 54065.36, entropy of 0.849, whereas the 6-class solution yields BIC value of 53607.492, with entropy of 0.767). As the BIC fit-index 'plateaus' there is no additional benefit in increasing the number of latent classes in the model. BIC and AIC logic dictate that the optimal model is one with the lowest value, as is evident from Figure 3, the six-class solution was selected as the optimal model.

3.2   Latent classes

Table 1 illustrates the unique characteristics for each of the six homogeneous latent classes, together with the average class probabilities and 'anecdotal labels' used to describe them. Class 5 is relatively small in size (n= 183), however this class proves very valuable in terms of the emerging trends in drug consumption patterns and reinforces anecdotal evidence (EMCDDA, 2002).

- Class 1, titled 'Clubbers' consists of 950 individuals (13.6%), characterised by a high probability of cannabis and ecstasy use (0.88 and 0.84 respectively) with alcohol and amphetamines (speed) the next most probable drugs of misuse (0.39 and 0.43). The 'Clubbers' are the class most likely to misuse hallucinogenic drugs such as LSD or magic mushrooms (16.8) and are unlikely to use the harder drugs, such as heroin and methadone (0.07 and 0.01).
- Class 2, entitled 'Novices' represents 8.3% of the total sample (n=577). Individuals in this class have a very high probability of cannabis use (0.93) and are also most likely to misuse volatile inhalants such as glue and/or solvents (0.10). There is a low probability of misuse of harder drugs such as amphetamines, cocaine or heroin (0.02, 0.03 and 0.08).
- Class 3, 'Polydrug users' (n = 1013, 14.5%) are those receiving treatment for multiple drugs of misuse, in this case three or more drugs. This class has a very high probability of misuse of heroin (0.97), with the next most probable drug of misuse being cannabis (0.65). Individuals classed as 'Polydrug users' are the most likely to abuse cocaine (0.29).
- The 999 individuals (14.3%) in class 4 have a probability of 1.0 of misusing methadone. These individuals have a high probability of heroin use (0.86) and benzodiazepines (0.36) in combination with methadone. The harder drugs dominate this class and there is a low probability of abuse of 'softer drugs' (ecstasy 0.05 and amphetamines 0.01) as well as alcohol (0.02).
- Class 5 is the smallest of all the classes representing 183 individuals (2.6%). Individuals in this 'Benzodiazepine misusers' class misuse

benzodiazepines, such as valium, dalmane and rohypnol (1.0). Included in this class are individuals who also misuse other opiates, such as codeine and distalgesics (0.40) or alcohol (0.25). Anecdotal evidence suggests the misuse of benzodiazepines as an emerging trend (EMCDDA, 2002), and thereby this evidence from 2000 warrants close investigation.

- Class 6 is the largest of all the latent classes with 3272 individuals, representing 46.8% of the sample. This class, entitled 'Heroin misusers' is dominated by the abuse of heroin (1.0). For members of this class, if more than one drug was being misused, the next most common drugs together with heroin were benzodiazepines (0.30) and cannabis (0.12).

The average class probabilities indicate the level of accurate classification within each of the latent classes (diagonal, Table 1). The class probabilities on the 'off-diagonal' represent misclassifications between each of the classes, for example, the misclassification between 'Methadone addicts' (Class 4) and 'Clubbers' (Class1) was 0.001.

Respondents were assigned to classes based on their probability of inclusion in that class; it was this conditional probability score that was correlated with covariates for the multinomial logistic regression. In this way, the probability of being included in a specific class was not constrained to be certain (1.0), as this would have introduced assignment error.

3.3   Multinomial logistic regression

Table 2: Multinomial logistic regression equations for the latent classes regressed on age (x) and early school leavers (y).

	Age $\alpha_j + \beta_j x_i,$	Age and Early School Leavers $\alpha_j + \beta_j x_i + \beta_i y_i$
Class 1 equation =	1.853 -0.129*(age)	0.094 -0.121*(age) + 0.840*(school)
Class 2 equation =	2.808 -0.18* (age)	1.133 -0.130*(age) + 0.207*(school)
Class 3 equation =	-0.339 -0.02*(age)	-0.651 -0.020*(age) +0.193*(school)
Class 4 equation =	-1.124 +0.0*(age)	-0.663 -0.006*(age) -0.204*(school)
Class 5 equation =	-4.703 +0.084*(age)	-5.594 +0.091*(age) +0.348*(school)

The estimated probabilities of class membership for a given age (Equation 2.2) equal

$$\pi_{i1} = \frac{\exp(1.853 - 0.129 x_i)}{1 + \exp(1.853 - 0.129 x_i) + \exp(2.808 - 0.18 x_i) + \exp(-0.339 - 0.02 x_i) + \exp(-1.124 + 0.0 x_i) + \exp(-4.703 + 0.084 x_i)}$$

$$\pi_{i2} = \frac{\exp(2.808 - 0.18x_i)}{1 + \exp(1.853 - 0.129x_i) + \exp(2.808 - 0.18x_i) + \exp(-0.339 - 0.02x_i) + \exp(-1.124 + 0.0x_i) + \exp(-4.703 + 0.084x_i)}$$

through to…

$$\pi_{i6} = \frac{1}{1 + \exp(1.853 - 0.129x_i) + \exp(2.808 - 0.18x_i) + \exp(-0.339 - 01.02x_i) + \exp(-1.124 + 0.0x_i) + \exp(-4.703 + 0.084x_i)}$$

The '1' term in each of the denominators and in the numerator of the π_{i6} represents $\exp(\alpha_{iJ} + \beta_{iJ})$ using $\alpha_{iJ} = \beta_{iJ} = 0$, where J is class 6. The six probabilities add to 1 as the numerators sum to the common denominator.

The 'Novices' class dominates the younger ages and almost disappears by the early thirties. The probability of membership in the 'Clubbers' class peaks at 20 years of age and declines slowly through the thirties. The 'Methadone' and 'Polydrug' classes remain very robust over the spectrum of ages. From the early-twenties through to the late-forties, individuals receiving treatment for drug misuse are most likely to be members of the 'heroin addicts' class. For older drug users, (40 years +), there is a sharp rise in the probability of membership of the 'Benzodiazepine misuser' class. Figure 4 illustrates the multinomial logistic regression paths for each of the classes regressed upon 'age' (continuous variable) graphically modelled within MSExcel.
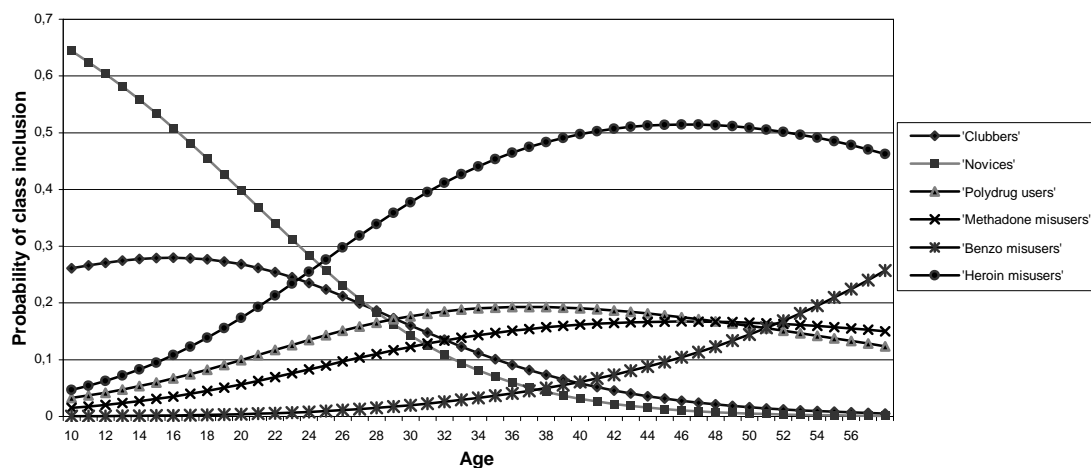

Figure 4: Six Latent Classes regressed on Drug Users Age.

Figure 5 illustrates the regression paths for each of the six latent classes regressed on drug user's age, and also regressed on those who left school before aged 15, and those who stayed in school after the age of 15 years. All the solid lines represent those who left school early. Individuals who stayed in school longer were more likely to be members of the 'Clubbers' class (recreational drugs use), while the probability of early school leavers being in the 'Novices' class declines sharply in the late teens. It is also apparent that early school leavers are

more likely to be members of the 'Heroin' class over all of the ages and less likely to be members of the 'Benzodiazepine misuser' class in later years.

It should be noted that the actual number of drug users in treatment decreases with age and although the probability of membership of the 'Benzodiazepine misusers' class increases with age, this may be as a function of a declining proportion in the other latent classes.
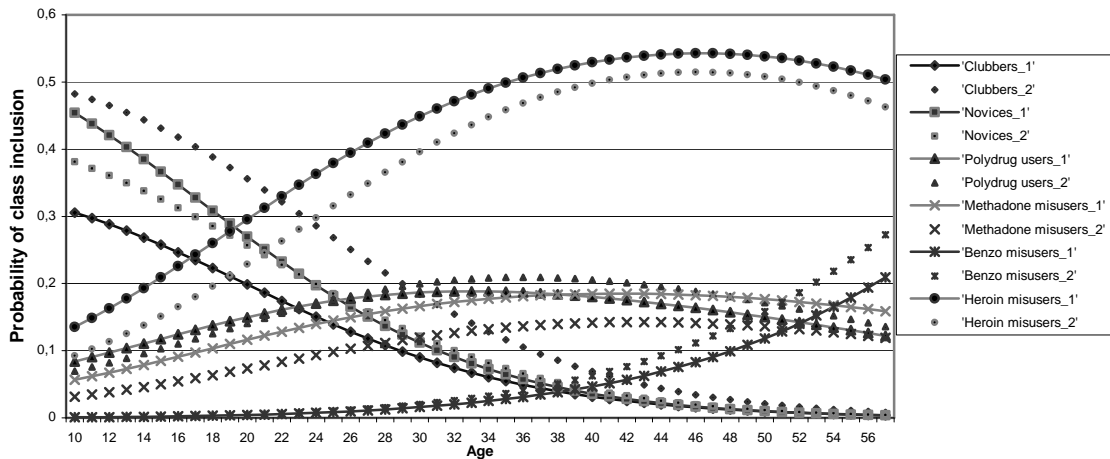

Figure 5: Six Latent Classes regressed on Drug User's Age, and those who left school before age 15 solid line_1) and those who stayed in school after age 15 years (broken line_2).

4   Conclusions

The current study utilised latent variable mixture models to examine profiles of drug misusers and incorporated these methodologies for the analysis of treated drug misuse in the Republic of Ireland

Using both the Bayesian and Akaike information criteria (BIC and AIC) a six-class solution was selected as the optimal number of latent classes 'best-fitting' the data.   From the subsequent unconditional latent class analyses, the characteristics of each of the classes were uncovered and each of the typical profiles was developed.

The diverse drug misusing behaviours of the 6994 individuals who received treatment for drug misuse were structured into six distinct and homogeneous classes (Clubbers n=950, Novices n=577, Polydrug users n=1013, Methadone misusers n=999, Benzodiazepine misusers n=183 and Heroin misusers n=3272).

Given the complex nature of problems associated with drug misuse, this research highlighted the need to avoid a single treatment modality for problem drug use, and showed that a range of treatment options may need to be considered for the unique characteristics of each class.

A series of conditional models was then examined in terms of the key covariates 1) age and 2) age and early school leavers. The parameter estimates

obtained from the multinomial logistic regression were graphically modelled with MSExcel and the respective functions described.

Stemming from a better understanding of the probability of an individual being in a given latent class dependent on age (continuous) and early school leavers (dichotomous), intervention strategies both reactive (health care policies) and proactive (social and educational initiatives) may be more fully evaluated.

References

[1] Agresti, A. (2002): *Categorical Data Analysis (2nd ed).* NY: Wiley and Sons.

[2] Department of Tourism, Sport and Recreation. (2001): *Building on Experience. National Drugs Strategy 2001-2008.* Dublin: The Stationery Office.

[3] EMCDDA (2002): European monitoring centre on drugs and drug addiction. *Annual Report on the State of the Drugs Problem in the European Union.* Luxembourg: Office for Official Publications of the European Communities.

[4] EMCDDA and Pompidou Group (2000): *Treatment Demand Indicator: Standard Protocol 2.0.* Lisbon: European Monitoring Centre for Drugs and Drug Addiction.

[5] Gossop, M., Marsden, J., and Stewart, D. (2001): *NTORS. After Five Years. The National Treatment Outcome Research Study.* London: National Addiction Centre.

[6] Hagenaars, J.A. and McCutcheon, A.L. (Eds.) (2002): *Applied Latent Class Analysis Models.* UK: Cambridge University Press.

[7] Hosmer, D.W. and Lemeshow, S. (2000): *Applied Logistic Regression.* New York: John Wiley and Sons.

[8] Muthén, L. and Muthén B. (2003): MPlus short courses. *Special Topics in Latent Variable Modelling Using Mplus.* Fulda, Germany.

[9] O'Brien, M., Kelleher, T., and Cahill, P. (2002): *Trends in Treated Drug Misuse in Health Boards 1996-2000.* Health Research Board, Dublin.

[10] Schwarz, G. (1978): Estimating the dimensions of a model. *Annals of Statistics*, **6**, 461-464.

[11] Simpson, D.D. and Sells, S.B. (Eds.). (1990): *Opioid Addiction and Treatment: A 12-Year Follow-up.* Malabar, Florida: Krieger.