

# Illicit Drug Use and Labour Market Achievement: Evidence from the UK\*

Ziggy MacDonald and Stephen Pudney  
Public Sector Economics Research Centre  
Department of Economics  
University of Leicester

June 1998

## Abstract

This study, using data from the British Crime Survey (BCS), examines the effect of drug use on occupational achievement. We start by attempting to overcome the identification problem that results from the limited set of drug use questions presented in the BCS. Taking this into account, and allowing for the endogeneity of drug use in equations for unemployment and labour market outcomes, we observe quite different relationships for soft and hard drug use. For soft drugs, there is a positive association with occupational achievement that diminishes with age. This relationship also holds for males but not for females. In contrast, we also find that past use of hard drugs significantly increases the likelihood of current unemployment, although it appears to be unrelated to occupational success, conditional on achieving employment.

*JEL classification:* C51; I12; J24

*Keywords:* Illicit drugs; labour market participation; Productivity

---

\*Material from Crown copyright records made available through the Home Office and the ESRC Data Archive has been used by permission of the Controller of Her Majesty's Stationery Office.

# 1 Introduction

The economic foundations of anti-drug policy rest on the belief that drug use imposes serious costs on individuals and on society as a whole. The latter external costs include an increased burden on publicly provided health care and the impact of acquisitive crime linked to drug use. They create a divergence between the marginal private costs of the individual decision maker (the drug user) and the marginal social costs borne by society as a whole, and give rise to a strong case for government intervention. However, in his original analysis of the drugs problem, Culyer (1973) highlighted particularly the negative impact of drug consumption on productivity. The primary concern of this traditional view is that reduced labour market experience of drug users will ultimately result in a lower aggregate level of human capital accumulation, tending to reduce overall productivity and hence living standards. This reduced labour productivity is usually seen to result from the negative health consequences of drug use which can lead to chronic absenteeism and frequent spells out of the labour market. There is considerable sociological and psychological research to suggest that drug users tend to be less socially conforming and have lower commitment to traditional indicators of social success, such as career progression (Kandel, 1984). It is also clear that drug users may give unfavourable signals to employers through their values and behaviour (Kandel *et al.*, 1995). Thus, assuming that workers receive the value of their marginal product as pay, then the reduced productivity level of drug users would manifest itself through lower wages.

In recent years, however, there have been a number of papers questioning this view of the relationship between illicit drug use and productivity. The essence of this work is the recognition of the possible simultaneity of drug use and wages, and the existence of unobserved heterogeneity, which raise questions about the direction of causality in a wage equation involving a measure of drug use as an explanatory variable. The endogeneity issue follows from conventional consumption-labour supply theory in which drug use is treated only as one form of consumption, determined optimally in response to the market wage and non-labour income. If one goes on to assume that an individual's wage is a negative function of illicit drug use, then causality between drug use and wages must be bi-directional. The related issue of heterogeneity derives from the fact that the unobserved attributes that affect wages could be the same characteristics that influence an individual's choice to take drugs. This will be true if the unobserved characteristic is a high rate of time preference which will cause individuals to select high paying jobs, but also make individuals much more likely to take drugs (Becker and Murphy, 1988). Sociological characteristics such as poor self-esteem and

social alienation would tend to work in the opposite direction.

Kaestner (1991) was the first to address these problems in the labour economics literature. Using data from the US National Longitudinal Survey of Youth (NLSY), he used the Heckman two-stage procedure to estimate a wage equation. The results of this study suggest that drug use does not have a negative impact on wages as previously expected. Kaestner finds that, if anything, increased frequency of illicit drug use (in this case cocaine or marijuana) tends to lead to higher wages, a result consistent across gender and age groups. Building on this work, Gill and Michaels (1992) and Register and Williams (1992), using the same data, find similar results. Gill and Michaels (1992) takes Kaestner's work further by modelling the effect of drug use on the likelihood of labour market participation. They estimate wage equations for users and non-users taking account of self-selection. They find a strong positive relationship between wages and the use of hard drugs, net of selectivity bias. Similar results are found by Register and Williams (1992), who use a continuous measure of drug use, exclude women from their sample, and use a different methodology for controlling the self-selection of individuals into drug use and the labour market. An update of these papers is presented in Kaestner (1994a, 1994b), which update previous cross-sectional estimates and provide longitudinal estimates using two waves of the NLSY. The cross sectional results are generally consistent with the previous studies, but the longitudinal estimates only provide partial support for the positive relationship between drug use and wages. In particular, he finds a wide range of wage effects that depend on the type of drug and individual: for example a positive relationship between cocaine use and wages for females, but a negative relationship between marijuana use and wages for males. Timing is also important: for example recent marijuana use is found to have a positive effect for males, but lifetime use a negative effect.

A major criticism of this work is the lack of explanation for the counter-intuitive results. Typically an explanation is offered in terms of some unobserved characteristic that is positively correlated with wages and illicit drug use (e.g. rate of time preference or an extrovert nature). Kandel *et al.* (1995) offer an alternative explanation, however. They point to the relative youth of the NLSY as the source of these results, suggesting that an allowance for a life-span perspective would help rationalise these findings. Kandel *et al.* suggest that the relationship between drug use and wages depends upon whether the worker is in early or later career stages. This hypothesis is partially motivated by Kaestner's (1994a, 1994b) slightly conflicting longitudinal results. The authors use a follow-up cohort of the NLSY who were last interviewed at age 34-35. The results suggest a positive relationship between drug use and wages in the early stages of an individual's career, but a negative relation-

ship later on in the career. Using the same data source, Burgess and Proper (1998) find a contradictory result in their analysis of the effects of early life behaviour that is considered anti-social or deviant (such as drug and alcohol consumption) and later life outcomes (such as productivity and household formation). Their main findings suggest that, particularly for men, adolescent alcohol and (soft) drug use has little or no effect on the earnings of men in their late twenties, a result that extends to earnings 10 years on. They do find, however, a negative relationship between adolescent heavy drug use and the earnings and labour market participation of young men.

In this paper we attempt to address these issues using data from the 1994 British Crime Survey (BCS). In particular, we estimate a joint model covering past and current drug use together with unemployment and occupational success. In order to consider the life-time perspective our model is estimated separately for a younger cohort (aged 16-29) and an older cohort (aged 30-59). Before developing our empirical model, we first discuss the BCS data set, and its advantages and shortcomings.

## 2 The Data

In contrast to the US, the UK undertakes very little monitoring of drug use at a national level, a problem typical of most European countries (Farrell *et al.*, 1994). Until recently the Home Office maintained an index of addicts (although only for users of opiates and cocaine), but this was closed in May 1997 with notifications being transferred to the regional drug misuse databases (Tregoning, 1997). The regional databases, although well developed, are limited in what they can monitor by way of drug use. Although they record information on drug use beyond opiates and cocaine, they only compile information on episodes of drug misuse that generate demand for treatment. In other words, the regional databases only provide information on the sub-set of drug users who present themselves to community-based agencies for problem drug misuse. Apart from these sources of information, and the occasional local survey, the only other major source of drug misuse information in the UK is the British Crime Survey (BCS).

### 2.1 The British Crime Survey

Information on illicit drug use was collected in the BCS for the first time in 1992, then in subsequent surveys in 1994 and 1996. The 1992 survey is generally considered not suitable for analysis and comparison with the 1994 survey (Ramsay and Percy, 1997) and the 1996 survey is not yet in the pub-

lic domain (although preliminary results have been published by the Home Office: see Ramsay & Spiller, 1997). Thus we use the 1994 wave of the BCS for our analysis. The sampling framework uses the Postcode Address File from which a sample size of approximately 14500 individuals was generated in 1994. The sampling process is designed to yield a representative sample of adults aged 16 and over living in England and Wales. The core sample is boosted with an additional representative sample of 2000 ethnic minority adults. In comparison with 1991 Census data and the 1993 General Household Survey, the 1994 BCS is known to under-represent men aged 20-29 and over-represent women aged 30-39. There is also a slight under-representation of working men, but on the whole the survey is considered to be representative of England and Wales. For more details on the sampling procedures see White and Malbon (1995). From the core and booster samples, approximately 10,000 eligible individuals completed the drug-use self-completion component of the survey. A summary of the findings for 1994 is given in Ramsay and Percy (1996).

## **2.2 Drug classification**

In the BCS, questions on drug-use are presented via an additional self-completion questionnaire, completed by respondents between 16 and 60 years of age. In 1994 the completion method changed from a paper-based to a computer-based system. The 1994 survey lists 13 of the most commonly abused drugs plus the bogus drug Semeron (put in the survey to test for false claiming - only one respondent claimed to have taken this drug). The survey also includes 3 catchall questions to capture those drugs not listed. A complete list of the drugs (including their popular alternative names) is provided Table A1 in the Appendix. Survey respondents were asked four questions about each of the drugs listed: had they heard of the drug, had they ever taken it, had they taken the drug in the past 12 months or in the past month. Although somewhat limited in scope, these questions provide us with some information about an individual's prevalence of past and recent drug use. The data for 1994 usage are summarised in Table I, together with summary information on employment. We have categorised the data by drug type, separating use of class A hard drugs (heroin, cocaine, LSD, etc.) from the use of class B and C soft drugs (mainly cannabis).

Table I  
 Summary of illicit drug use and employment (%)  
 (standard errors in parentheses)

	All	Males	Females	Age 16-29	Age 30-59	In work	Out of work
<i>Class A drugs</i>							
Never used	91.0 (0.339)	89.1 (0.506)	93.2 (0.436)	84.6 (0.815)	93.4 (0.345)	91.5 (0.342)	84.1 (1.639)
Only used in past	7.3 (0.309)	8.9 (0.462)	5.5 (0.397)	10.9 (0.703)	6.0 (0.331)	7.1 (0.315)	11.0 (1.406)
Recently used	1.6 (0.151)	2.0 (0.227)	1.2 (0.192)	4.5 (0.468)	0.6 (0.104)	1.4 (0.145)	4.8 (0.961)
<i>Class B/C drugs</i>							
Never used	75.1 (0.513)	72.1 (0.727)	78.5 (0.714)	65.5 (0.107)	78.7 (0.570)	75.8 (0.526)	64.9 (2.141)
Only used in past	16.5 (0.439)	18.0 (0.623)	14.7 (0.615)	15.7 (0.821)	16.8 (0.520)	16.6 (0.457)	14.9 (1.595)
Recently used	8.5 (0.330)	9.9 (0.484)	6.8 (0.439)	18.8 (0.883)	4.5 (0.290)	7.6 (0.325)	20.3 (1.804)
Unemployed	7.0 (0.302)	10.0 (0.486)	3.6 (0.626)	8.4 (0.626)	6.5 (0.343)	-	-
Observations	7188	3801	3317	1960	5158	6620	498

It is clear from Table I that there is a far greater use of soft than hard drugs, with over 91% of the sample never having tried class A drugs. Also, Table I suggests that the young and the currently unemployed have the highest relative frequency of recent drug use, both hard and soft. In particular, a greater percentage of the 16-29 age group have recently taken drugs than have any other group in the past. Similarly, those currently out of work have the highest relative frequency of recent drug use, a fifth of this group having recently taken soft drugs.

### 2.3 Employment information

Previous analysis of the relationship between drug use and productivity has made use of individual data on earnings or wages and, in some cases, drug use at an intensive margin rather than the extensive margin (Kaestner, 1991, 1994a, 1994b; Register & Williams, 1992). The BCS, however, does not provide information on individual earnings or wages. Rather, it reports total

household income defined in terms of income bands. This means that for the purpose of our analysis we use occupational class as a proxy for labour market productivity. In some respects this may be preferable to using wages as a measure of productivity given that for many individuals their productivity is not reflected by the wage they receive. This is particularly true for young workers who take drugs, where the employer cannot observe the use of drugs directly and has no prior information on productivity (e.g. previous participation and achievement) to base this judgement on. The BCS offers a standard definition of occupational class, ranging from professional/managerial, through to partly skilled and unskilled. This classification can be considered as a reasonable indicator of labour market achievement and hence productivity. For our analysis we work with the four occupational classes summarised in Table II.

Table II  
Summary of occupational classes (%)  
(Standard errors in parentheses)

	All	Age 16-29	Age 30-59	Used class A	Never used class A	Used class B/C	Never used class B/C
Managerial/ Professional	35.8 (0.568)	24.8 (0.976)	40.0 (0.682)	37.6 (1.917)	35.6 (0.595)	40.2 (1.165)	34.3 (0.650)
Skilled	39.4 (0.579)	47.7 (1.128)	36.4 (0.670)	32.7 (1.857)	40.1 (0.609)	35.8 (1.139)	40.6 (0.672)
Partly Skilled	13.5 (0.405)	15.2 (0.811)	12.8 (0.465)	13.5 (1.351)	13.5 (0.424)	10.8 (0.738)	14.4 (0.480)
Unskilled	4.3 (0.241)	4.2 (0.542)	4.4 (0.285)	3.9 (0.768)	4.4 (0.254)	3.3 (0.422)	4.7 (0.289)
Unemployed	7.0 (0.302)	8.0 (0.626)	6.5 (0.343)	12.4 (1.303)	6.5 (0.306)	9.9 (0.708)	6.0 (0.326)
Observations	7118	1960	5158	639	6479	1774	5344

The figures in Table II clearly suggest that a relatively higher frequency of drug users are unemployed compared to non-users, particularly those who have used class A drugs (either in the past or currently). However, when drug users are employed, there is some suggestion of a positive association with labour market achievement. For instance, those in the managerial/professional class have a higher proportion of users than non-users. Although this difference is only slight for hard drugs, it is particularly noticeable for soft class B/C drugs. This observation is further confused when one considers the experience of younger and older respondents separately. As we

have seen above, a greater proportion of younger respondents report drug taking, and from Table II, younger respondents are proportionately more represented in the unemployment category. Unlike the experience of drug users in general, however, younger respondents have a much lower representation in the highest occupational class, although they are well represented in the skilled class.

### 3 A Model of Drug Use and Labour Market Achievement

Consider an individual, interviewed at a particular date. Divide his or her life into two periods: a past period finishing 12 months before the interview date; and a current period consisting of the 12 months leading up to the interview. We are interested in four events:

- (i) drug use during the past period (binary indicator  $d_1$ );
- (ii) drug use during the current period (binary indicator  $d_2$ );
- (iii) unemployment at the survey date (binary indicator  $u$ );
- (iv) occupational achievement (if in employment) at the survey date (categorical indicator  $a$ ).

A particular concern is the influence that past and current drug use might have on current labour market achievement. If all four events (i)-(iv) were directly observable, we would wish to estimate the following two-part probability structure:

*Currently unemployed*

$$\Pr(d_1, d_2, u = 1|x) = \Pr(d_1|x) \Pr(d_2, u = 1|d_1, x) \quad (1)$$

*Currently employed*

$$\Pr(d_1, d_2, u = 0, a|x) = \Pr(d_1|x) \Pr(d_2, u = 0, a|d_1, x) \quad (2)$$

where  $x$  is a vector of exogenous explanatory variables.

#### 3.1 Identification and survey design

Unfortunately, there is a serious observational problem stemming from the design of the questionnaire used in the 1994 BCS (and other European surveys). The respondent is asked only whether or not he or she has ever used



drugs, and, if so, whether or not within the last year.<sup>1</sup> Thus, concentrating only on the drug use variables  $d_1$  and  $d_2$ , we are able to identify three rather than four possible cases:

*No use ever*

$$\Pi_1(x) = (1 - P_1(x)) P_{00}(x) \quad (3)$$

*Early use, but no current use*

$$\Pi_2(x) = P_1(x) P_{10}(x) \quad (4)$$

*Current use*

$$\Pi_3(x) = (1 - P_1(x)) P_{01}(x) + P_1(x) P_{11}(x) \quad (5)$$

where  $\Pi_1(x)$ ,  $\Pi_2(x)$  and  $\Pi_3(x)$  are the conditional probabilities of the three observable states,  $P_1(x)$  is  $\Pr(d_1 = 1|x)$ , and the four double-subscripted probabilities are  $P_{ij}(x) = \Pr(d_2 = j|d_1 = i, x)$ , where  $i, j = 0$  or  $1$ .

### 3.1.1 The identification problem

The observable probabilities (3)-(5) sum identically to 1, so there are only two independent quantities,  $\Pi_1(x)$ , and  $\Pi_2(x)$ , from which to derive estimates of the structural probabilities  $P_1$ ,  $P_{00}$  and  $P_{10}$ . Consequently, these are not identifiable. Specifically, for any arbitrary function  $P_1^*(x) \neq P_1(x)$  mapping the domain of  $x$  into the admissible interval  $[P_1(x)P_{10}(x), 1 - P_{00}(x)(1 - P_1(x))]$ , it is possible to construct functions  $P_{00}^*(x) = \Pi_1(x)/(1 - P_1^*(x))$  and  $P_{10}^*(x) = \Pi_2(x)/P_1^*(x)$  such that (3)-(4) remain satisfied when  $P_1$ ,  $P_{00}$  and  $P_{10}$  are replaced by  $P_1^*$ ,  $P_{00}^*$  and  $P_{10}^*$ . The structural probabilities are therefore not separately identifiable in this general nonparametric sense. There is no doubt that this widely-used form of the survey questions is a constraint on the degree of robustness that can be achieved, and should be reconsidered by survey designers.

### 3.1.2 Identification through restrictions on functional form

However, to conclude from this general analysis that it is impossible to produce meaningful estimates of the structure is unduly pessimistic, since it

---

<sup>1</sup>Or month; however, we only present results in this paper for the 1 year recall question, since the number of positive responses is considerably smaller for the 1 month recall period. As a result, we are not excluding from our analysis those who use drugs only very occasionally.

assumes that there is no necessary relationship between the pattern of current drug use for those who were past users ( $P_{10}$ ,  $P_{11}$ ) and those who were not ( $P_{00}$ ,  $P_{01}$ ). In fact, it is perfectly reasonable to follow standard practice and capture the distinction between the two groups by including a simple lag effect in a conventional model of current drug use. In general terms, this gives the following structure:

$$\Pi_1(x) = [1 - P_1(x)] G(t(x, \beta)) \quad (6)$$

$$\Pi_2(x) = P_1(x) G(t(x, \beta) - \delta) \quad (7)$$

where  $t(\cdot)$  is some known (usually bilinear) function of  $x$  and an unknown parameter vector  $\beta$ , and  $G(\cdot)$  is a known distribution function (usually the normal or logistic). In this case, consider some choice  $\beta^* \neq \beta$ ; then a corresponding function  $P_1^*(x) = 1 - \Pi_1(x)/G(t(x, \beta^*))$  is implied by (6). For identification to fail, it must then be possible to find a constant  $\delta^*$  such that  $P_1(x) G(t(x, \beta) - \delta) = P_1^*(x) G(t(x, \beta^*) - \delta^*)$  is satisfied for all  $x$ . In general, since  $P_1$ ,  $G$  and  $t$  are arbitrary, there is no reason why this should be so for any values other than  $\{\beta^*, \delta^*\} = \{\beta, \delta\}$ . Note, however, that the structure is only marginally identified - for example, if we specify  $G$  as an unknown cdf to be estimated non-parametrically, it is always possible to find multiple choices of  $\{P_1, G, \beta, \delta\}$  that satisfy (6)-(7).

### 3.1.3 Identification through restrictions on drug use transitions

A more drastic solution to the identification problem is to restrict the functions  $P_{00}$  and  $P_{01}$  to be equal to 1 and 0 respectively, implying that current drug use is essentially impossible unless early use has occurred. This is a realistic assumption for the older cohorts since a common finding in studies of drug users is that individuals tend to mature out of drug use around the ages of 28 to 35 (Gill & Michaels 1991, Johnston *et al.* 1988, Kandel 1980, Labouvie 1996, MacDonald 1997, Ramsay and Percy 1996). Therefore it is extremely unlikely, for example, that a 30-year old with no previous experience of drugs will suddenly start using drugs. Restricting the analysis to the older cohorts has some advantages in any case, since it allows more time for the long-term impact of past drug use to become apparent. In this case, the structure (3)-(4) reduces to the following:

*No use ever*

$$\Pi_1(x) = 1 - P_1(x) \quad (8)$$

*Past use, but no current use*

$$\Pi_2(x) = P_1(x)P_{10}(x) \quad (9)$$

The remaining functions  $P_1$  and  $P_{10}$  are clearly identifiable now, since  $P_1(\cdot)$  can be constructed as  $1 - \Pi_1(\cdot)$  and  $P_{10}(\cdot)$  as  $\Pi_2(\cdot)/(1 - \Pi_1(\cdot))$ .

### 3.2 Model specification and estimation

Consider first the determination of past drug use. Define a latent variable  $d_1^*$  representing an individual's past propensity to consume drugs. This drives the observed indicator of actual drug use,  $d_1$ , through the usual probit mechanism:

$$d_1^* = x_1\beta_1 + \varepsilon_1 \quad (10)$$

$$d_1 = \Upsilon(d_1^* > 0) \quad (11)$$

where  $\Upsilon(\cdot)$  is the indicator function, equal to 1 if the argument is true and 0 otherwise;  $x_1$  is a row vector of personal and demographic attributes,  $\beta_1$  is the corresponding vector of parameters, and  $\varepsilon_1$  is a random error distributed as  $N(0,1)$  conditional on  $x$ .

The second stage of the model determines current drug use, experience of unemployment and occupational success jointly, but conditional on early drug use. Thus we have a system of three latent variables, assumed to be generated by the following multivariate regression structure:

$$d_2^* = x_2\beta_2 + d_1\delta_2 + \varepsilon_2 \quad (12)$$

$$u^* = x_3\beta_3 + d_1\delta_3 + \varepsilon_3 \quad (13)$$

$$a^* = x_4\beta_4 + d_1\delta_4 + \varepsilon_4 \quad (14)$$

where  $x_2...x_4$  are row vectors of personal and demographic attributes,  $\beta_2... \beta_4$  are the corresponding vectors of parameters, and  $\varepsilon_2... \varepsilon_4$  are errors with a trivariate normal distribution with zero means, unit variances and unrestricted correlations, conditional on  $x = \{x_1, x_2, x_3\}$  and  $d_1$ .

The observable counterparts of these latent variables are the binary indicators of current drug use and unemployment  $d_2$  and  $u$ , and an ordered categorical indicator of occupational achievement,  $a$ , ranging from unskilled ( $a = 1$ ) to professional/managerial ( $a = 4$ ). If we adopt the usual probit and ordered probit structures for these, the latent variables are assumed to generate the observed states by means of the following relationships:

$$d_2 = \Upsilon(d_2^* > 0) \quad (15)$$

$$u = \Upsilon(u^* > 0) \quad (16)$$

$$a = r\Upsilon(C_{r-1} \square a^* < C_r), \quad r = 1..4 \quad (17)$$

where  $C_0 = -\infty$ ,  $C_4 = +\infty$  and  $C_1..C_3$  are unknown threshold parameters.

From this structure, it is possible to derive the conditional probabilities of the 15 possible observational outcomes. These probabilities are tedious, and are relegated to appendix 2. In general they require the evaluation of trivariate normal probabilities, so the computational difficulty of maximum likelihood estimation is significant. As it stands, the structure is formally identified because it incorporates sufficient functional form restrictions. However, there may be practical problems in obtaining good estimates in cases like this where identification is fragile, and we may wish to go further and impose the restriction  $P_{00} = 1$ . In this case the lag parameter  $\delta_2$  can be dropped and  $d_2$  is treated as conditionally nonstochastic ( $= 0$ ) whenever  $d_1 = 0$ . This simplifies the computational problem somewhat, since some of the trivariate normal probabilities are reduced to bivariate ones.

## 4 Results

We have estimated the model described above in a number of ways. We present separate estimates for class A and class B/C drugs, since these represent a widely accepted distinction between hard and soft drugs. We also compute separate estimates for young and old cohorts. In addition, we estimate the full model as set out above, and the restricted version with  $P_{00}$  restricted to be unity; these are respectively referred to as the unrestricted and restricted estimates. For a description of all the variables used in this analysis and their descriptive statistics see Table A2 in Appendix 1. We begin by looking at the results for the older cohort, and consider first the impact of the identification problem generated by the survey design.

The past drug use component of the model (10)-(11) is specified very simply, reflecting work presented in earlier studies (Sickles and Taubman, 1991; MacDonald, 1997). Only basic demographic variables (age, education, ethnicity and gender) are included, while other current socio-economic variables that cannot be realistically regarded as exogenous determinants of early drug use are excluded. However, in line with previous studies (Sickles and Taubman, 1991) we have included a lifestyle variable to capture religious practice as this is likely to be fairly constant over time. The current drug use component is also specified as a probit, but with an expanded set of covariates

describing the current demographic nature of the individual and his or her household and also the lag effect of past drug use<sup>2</sup>. Similar sets of variables (differing mainly in the increased detail of the educational variables) are used as covariates in the binary probit for unemployment and the ordered probit for occupational achievement.

#### **4.1 The 30-59 cohort**

The cohort of people aged 30-59 at the time of the BCS in 1994 was born in the period 1935-64 and were therefore of school-leaving age around 1951-80, which was before the use of drugs in adolescence had become widespread, but was a period in which drug use became significant among (particularly middle-class) young adults.

We begin by examining the impact of the identification problem. Tables IIIa-d give the estimated coefficients separately for class A and B/C drugs, using both the unrestricted model (10)-(15) and the restricted model with  $P_{00}$  forced to be equal to 1.

---

<sup>2</sup>Unlike some recent studies (Ramsay and Percy 1996, Ramsay and Spiller 1997) we have not included alcohol abuse in the current drug use model. This is simply because drug use and alcohol abuse are joint outcomes of a given lifestyle driven by some unobserved process.

Table IIIa  
The probability of past drug use: estimates for 30-59 cohort  
(standard errors in parentheses)

Coefficient ( $\beta_1$ )	Class A drugs		Class B/C drugs	
	unrestricted	restricted	unrestricted	restricted
MALE	0.225*** (0.063)	0.246*** (0.063)	0.121*** (0.044)	0.169*** (0.042)
AGE	-0.351*** (0.045)	-0.360*** (0.044)	-0.227*** (0.031)	-0.309*** (0.027)
DEGREE	0.482*** (0.099)	0.442*** (0.100)	0.580*** (0.072)	0.580*** (0.066)
NON-DEGREE	0.147** (0.087)	0.083 (0.087)	0.194*** (0.059)	0.156*** (0.055)
BLACK	-0.147 (0.127)	-0.098 (0.123)	0.060 (0.087)	0.204*** (0.080)
ASIAN	- -	- -	-0.769*** (0.129)	-0.757*** (0.117)
OTHER	-0.037 (0.196)	-0.013 (0.205)	-0.260** (0.162)	-0.226* (0.147)
CHURCH	-0.274*** (0.081)	-0.289*** (0.081)	-0.171*** (0.054)	-0.266*** (0.051)
CONSTANT	-0.393** (0.205)	-0.276* (0.206)	-0.256** (0.151)	0.249** (0.132)
n	4868	4868	5158	5158

Note: \*\*\* = significant at 1% level, \*\* = significant at 5% level, \* = significant at 10% level

Table IIIb (continued over)  
The probability of current drug use: estimates for 30-59 cohort  
(standard errors in parentheses)

Coefficient ( $\beta_2$ )	Class A drugs		Class B/C drugs	
	unrestricted	restricted	unrestricted	restricted
MALE	0.455* (0.298)	0.674** (0.386)	0.284*** (0.078)	0.216*** (0.098)
AGE	-0.333** (0.189)	-0.188 (0.258)	-0.368*** (0.054)	-0.242*** (0.068)
INCITY	0.320 (0.206)	0.398 (0.316)	0.123** (0.073)	0.272*** (0.094)
DEGREE	0.090 (0.316)	-0.355* (0.400)	0.368*** (0.121)	-0.196 (0.148)
NON-DEGREE	-0.317 (0.264)	-0.765** (0.368)	0.036 (0.095)	-0.197* (0.133)
BLACK	0.025 (0.336)	0.189 (0.660)	0.220** (0.120)	0.176 (0.154)
ASIAN	- -	- -	-0.327* (0.216)	0.410 (0.348)
OTHER	0.268 (0.625)	0.585 (0.842)	-0.043 (0.251)	0.241 (0.371)
FAMTYP1	0.835* (0.626)	1.595** (0.917)	-0.013 (0.151)	0.125 (0.194)
FAMTYP2	0.407 (0.624)	0.719 (0.922)	-0.008 (0.131)	0.024 (0.171)
FAMTYP3	1.185 (0.766)	2.380* (0.994)	-0.205 (0.183)	0.009 (0.247)
FAMTYP4	0.799* (0.542)	2.140*** (0.833)	-0.024 (0.170)	0.038 (0.216)
FAMTYP5	0.466 (0.550)	0.468 (0.808)	-0.112 (0.121)	-0.112 (0.157)
MARRIED	-0.375* (0.282)	0.397 (0.416)	-0.674*** (0.106)	-0.517*** (0.131)

Note: \*\*\* = significant at 1% level, \*\* = significant at 5% level, \* = significant at 10% level

Table IIIb (continued)  
The probability of current drug use: estimates for 30-59 cohort  
(standard errors in parentheses)

Coefficient ( $\beta_2$ )	Class A drugs		Class B/C drugs	
	unrestricted	restricted	unrestricted	restricted
CHURCH	-0.201 (0.353)	-0.088 (0.528)	-0.367*** (0.101)	-0.246** (0.132)
CONSTANT	-1.976*** (0.869)	-2.228 (1.450)	0.039 (0.256)	0.293 (0.327)
Past use ( $\hat{\delta}_2$ )	-2.078 (202.91)	- -	-1.447 (2.134)	- -
n	4868	4868	5158	5158

Note: \*\*\* = significant at 1% level, \*\* = significant at 5% level, \* = significant at 10% level



Table IIIc (continued over)  
The probability of current unemployment: estimates for 30-59 cohort  
(standard errors in parentheses)

Coefficient ( $\beta_3$ )	Class A drugs		Class B/C drugs	
	unrestricted	restricted	unrestricted	restricted
MALE	0.732*** (0.089)	0.731*** (0.089)	0.751*** (0.088)	0.742*** (0.088)
AGE	0.071** (0.042)	0.071** (0.042)	0.061* (0.041)	0.072** (0.042)
INCITY	0.174*** (0.067)	0.173*** (0.067)	0.175*** (0.065)	0.180*** (0.065)
EDU1	-0.839*** (0.114)	-0.839*** (0.114)	-0.818*** (0.109)	-0.842*** (0.109)
EDU2	-0.664*** (0.110)	-0.664*** (0.110)	-0.693*** (0.109)	-0.697*** (0.109)
EDU3	-0.543*** (0.115)	-0.543*** (0.115)	-0.561*** (0.113)	-0.566*** (0.113)
EDU4	-0.395*** (0.090)	-0.393*** (0.090)	-0.406*** (0.087)	-0.413*** (0.087)
EDU5	-0.431*** (0.137)	-0.430*** (0.137)	-0.434*** (0.134)	-0.440*** (0.134)
EDU6	-0.144 (0.125)	-0.142 (0.125)	-0.164 (0.118)	-0.169* (0.119)
BLACK	0.344*** (0.108)	0.343*** (0.108)	0.332*** (0.107)	0.329*** (0.107)
ASIAN	-	-	0.192* (0.127)	0.216** (0.128)
OTHER	0.061 (0.244)	0.058 (0.244)	0.037 (0.249)	0.055 (0.254)
FAMTYP1	0.129 (0.138)	0.124 (0.138)	0.150 (0.132)	0.154 (0.133)

Note: \*\*\* = significant at 1% level, \*\* = significant at 5% level, \* = significant at 10% level

Table IIIc (continued)  
The probability of current unemployment: estimates for 30-59 cohort  
(standard errors in parentheses)

Coefficient ( $\beta_3$ )	Class A drugs		Class B/C drugs	
	unrestricted	restricted	unrestricted	restricted
FAMTYP2	-0.256** (0.123)	-0.258** (0.123)	-0.230** (0.117)	-0.234** (0.117)
FAMTYP3	-0.346*** (0.151)	-0.352*** (0.151)	-0.352*** (0.144)	-0.351*** (0.145)
FAMTYP4	0.287** (0.163)	0.283** (0.163)	0.342*** (0.157)	0.345*** (0.158)
FAMTYP5	-0.219** (0.113)	-0.220** (0.113)	-0.201** (0.106)	-0.204** (0.106)
MARRIED	-0.468*** (0.102)	-0.468*** (0.102)	-0.448*** (0.100)	-0.431*** (0.101)
CONSTANT	-1.704*** (0.221)	-1.702*** (0.221)	-1.672*** (0.213)	-1.750*** (0.218)
Past use ( $\hat{\delta}_3$ )	0.226** (0.119)	0.212** (0.118)	-0.057 (0.089)	0.098 (0.076)
n	4868	4868	5158	5158

Note: \*\*\* = significant at 1% level, \*\* = significant at 5% level, \* = significant at 10% level

Table IIIId (continued over)  
Occupational achievement: estimates for 30-59 cohort  
(standard errors in parentheses)

Coefficient ( $\beta_4$ )	Class A drugs		Class B/C drugs	
	unrestricted	restricted	unrestricted	restricted
AGE	0.079*	0.079*	0.065*	0.066*
	(0.026)	(0.026)	(0.025)	(0.025)
INCITY	-0.130*	-0.131*	-0.125*	-0.126*
	(0.038)	(0.037)	(0.036)	(0.036)
EDU1	2.317*	2.317*	2.236*	2.234*
	(0.068)	(0.068)	(0.065)	(0.066)
EDU2	1.563*	1.564*	1.536*	1.536*
	(0.062)	(0.061)	(0.061)	(0.062)
EDU3	1.078*	1.078*	1.041*	1.040*
	(0.070)	(0.070)	(0.068)	(0.069)
EDU4	0.724*	0.725*	0.705*	0.704*
	(0.054)	(0.054)	(0.053)	(0.053)
EDU5	0.367*	0.366*	0.345*	0.343*
	(0.075)	(0.075)	(0.075)	(0.075)
EDU6	0.421*	0.421*	0.400*	0.399*
	(0.079)	(0.079)	(0.075)	(0.075)
BLACK	-0.038	-0.038	-0.039	-0.039
	(0.075)	(0.075)	(0.073)	(0.073)
ASIAN	-	-	0.095	0.096
			(0.078)	(0.078)
OTHER	0.088	0.089	0.097	0.095
	(0.132)	(0.132)	(0.132)	(0.133)
FAMTYP1	0.255*	0.255*	0.252	0.253*
	(0.087)	(0.087)	(0.085)	(0.085)
FAMTYP2	0.162*	0.162*	0.169	0.169*
	(0.061)	(0.061)	(0.059)	(0.059)

Note: \*\*\* = signi cant at 1% level, \*\* = signi cant at 5% level, \* = signi cant at 10% level

Table IIIId (continued)  
Occupational achievement: estimates for 30-59 cohort  
(standard errors in parentheses)

Coefficient ( $\beta_4$ )	Class A drugs		Class B/C drugs	
	unrestricted	restricted	unrestricted	restricted
FAMTYP3	0.045 (0.074)	0.046 (0.074)	0.049 (0.070)	0.049 (0.071)
FAMTYP4	-0.099 (0.103)	-0.098 (0.104)	-0.118 (0.102)	-0.118 (0.102)
FAMTYP5	0.061 (0.057)	0.062 (0.057)	0.064 (0.054)	0.065 (0.054)
SNGLMALE	-0.232* (0.083)	-0.234** (0.083)	-0.244* (0.079)	-0.246* (0.079)
SNGLFMAL	-0.197* (0.071)	-0.198** (0.071)	-0.198* (0.071)	-0.201* (0.071)
MARRFMAL	-0.269* (0.059)	-0.268* (0.059)	-0.278* (0.057)	-0.279* (0.057)
Past use ( $\hat{\delta}_4$ )	-0.003 (0.078)	0.012 (0.080)	0.071* (0.048)	0.064* (0.045)
C <sub>1</sub>	-0.850* (0.149)	-0.845* (0.149)	-0.922* (0.145)	-0.917* (0.145)
C <sub>2</sub>	0.066 (0.138)	0.070 (0.138)	-0.006 (0.135)	0.000 (0.135)
C <sub>3</sub>	1.431* (0.131)	1.435* (0.131)	1.358 (0.128)	1.364* (0.129)
n	4868	4868	5158	5158

Note: \*\*\* = signi cant at 1% level, \*\* = signi cant at 5% level, \* = signi cant at 10% level

The first observation we can make about the results for the older cohort is that the restricted model (with  $P_{00}$  forced to be equal to 1) produces very similar results to the unrestricted model. The parameter estimates are of similar magnitude and, in the majority of cases, the same covariates are signi cant. For past drug use, we find that being male and having a degree or non-degree quali cation is signi cantly associated with increased drug use compared to the base (female, no formal quali cations, white, and non-church going). These findings are consistent for both class A and class B/C drugs, regardless of which model is estimated (restricted or not). Looking at negative associations, age and religious attendance signi cantly reduce the

likelihood of early drug use. We also found a significant negative association between individuals of Asian origin and early class B/C drug use.

As expected, the main differences in results can be seen for current drug use, although these differences are only slight. In particular, compared to the unrestricted model, the restricted model produces more significant explanatory variables for class A drugs but fewer for class B/C drugs. For class A drugs, where the significance changes there is no change in sign, but this is not the case for class B/C drugs. The estimated models have nothing to say about the process of transition from past to current drug use, owing to the fragile identification of the unrestricted model, as can be seen from the huge standard errors for  $\delta_2$ .

In the probit model for current unemployment, there is very little difference in the signs, magnitude or significance of explanatory variables across model specification and drug classes. We found that being male, older, located in an inner city area, and being a lone parent all have a significant positive impact on the likelihood of current unemployment. There is a significant negative effect of the level of education, living in an-all adult household or nuclear family, and being married. Although in general the results for the current unemployment model are consistent between drug types, the estimated impact of past drug use differs by drug class. Past use of class A hard drugs has a significant positive impact on the probability of current unemployment whereas there is no significant association with past use of soft drugs. Note that the approach used to resolve the identification problem makes no difference to the estimate of the lagged effect of drug use either here or in the occupational achievement equation:  $\hat{\delta}_3$  and  $\hat{\delta}_4$  are well-determined and robust.

For the occupational achievement equation, we also see little difference between the coefficient estimates of the restricted and unrestricted models. Regardless of model or drug type, we found that, conditional on being employed, older males with higher education fare best in the labour market, particularly if they are married. For this older cohort, however, we found no significant association between past use of hard drugs and labour market achievement, but a weak positive association with past use of soft drugs (a coefficient of less than 0.1, significant only at the 10% level).

The relationship between current drug use and unemployment or occupational success is given by the estimated correlation coefficients  $\hat{\rho}_{23}$  and  $\hat{\rho}_{24}$  respectively. The latter is never statistically significant, so there is no evidence of any current effect of hard or soft drug use on occupational success for those who are in work. The estimated impact of current drug use on the probability of unemployment is significant and positive ( $\hat{\rho}_{23} \approx 0.2$ ) for soft drug use, but insignificant for the hard drug version of the model.

## 4.2 The 16-29 cohort

The cohort of people aged 16-29 at the time of the BCS in 1994 were born in the period 1965-78 and were therefore of school-leaving age around 1981-94, when the use of drugs in adolescence had become widespread. For the younger cohort it makes little sense to impose the restriction  $P_{01} = 0$  on the drug use transition as the time period is much shorter, and it is quite possible that initiation into drugs may occur after age 16. Therefore we only estimate the unrestricted model for the 16-29 cohort. Unfortunately the identification problem proved insuperable for the hard drugs variant of the model, where the number of positive responses is necessarily small. As in all other cases, the lag parameter  $\delta_2$  was essentially unidentifiable, evidenced by a huge standard error. However, other lagged effects ( $\delta_3$  and  $\delta_4$ ) were well-determined. There were no obvious identification problems with the soft drugs variant of the model. We limit the presentation of results to the significant qualitative influences on current unemployment and labour market achievement, shown in Table IV below. The full results for the younger cohort are given in Tables A3i-iv in Appendix 3<sup>3</sup>.

The results in Table IV are consistent with other findings. We find that males, blacks, lone parents and single adults are more likely to be currently unemployed. The probability of current unemployment diminishes with level of educational attainment and also with living in an all-adult household. For those in work, labour market achievement is positively related with age, educational attainment, Asian origin and all household types except lone parents, who are significantly less likely to do well in the labour market in the soft drug variant of the model. These results are broadly consistent between drug class models, except for the effects of past drug use. In the soft drug version of the model, past drug use increases the probability of current unemployment but, for those in work, also increases the expected degree of occupational success. For the hard drugs version of the model, there is no significant effect of past drug use in this younger cohort.

The estimated correlation  $\hat{\rho}_{24}$  is statistically insignificant for both classes of drugs, implying no correlation between current drug use and occupational success for those in work, conditional on other characteristics. However, there is a significant association between current drug use and unemployment. For both soft drugs ( $\hat{\rho}_{23} = 0.32$ ) and hard drugs ( $\hat{\rho}_{23} = 0.45$ ), there is a significant increase in the probability of unemployment for drug users relative to non-users.

---

<sup>3</sup>We also estimated models separately for males and females, and for the whole sample. These are available from the authors.

Table IV

Significant associations with unemployment and labour market achievement

<b>Current Unemployment</b>		
	<i>Class A</i>	<i>Class B/C</i>
positive effect	male, inner-city, black, single adult household, lone parent, nuclear household	male, inner-city, black, single adult household, lone parent, past drug use
negative effect	all education levels, 2 or 3 adult households	all education levels, 3 adult household
<b>Labour market achievement</b>		
	<i>Class A</i>	<i>Class B/C</i>
positive effect	age, all education levels, Asian, all household types except lone parent, single female	age, all education levels, Asian, all household types except lone parent, female (single or married), past drug use
negative effect	black	inner-city, black, lone parent

## 5 Conclusion

The results presented in this paper are the first of their kind to be produced using UK data. We have highlighted a serious identification problem for statistical analysis of the relationship between drug use and labour market outcomes, stemming from the design of the drug use questions used in typical European surveys. However, we have developed alternative ways of overcoming this difficulty and have implemented them successfully.

Within the constraints imposed by the design of the British Crime Survey, the results presented here are consistent with the findings of recent US studies. We have looked at two dimensions of labour market success: unemployment and occupational achievement. For the older (30-59) cohort, in which the long-run effects of drug use are expected to be seen, we find that a history of hard (class A) drug use is significantly associated with unemployment. In contrast, there is no significant evidence of an effect of past use of soft (class B/C) drugs on the current incidence of unemployment, although there is a weak correlation with current use. Thus, any adverse effects of soft drugs appear not to be large or permanent. For the younger (16-29) cohort we find that the use of hard and soft drugs are much more strongly associated

with unemployment, although it is not possible to estimate the longer-term impact for this cohort.

For members of the older cohort who are in work, there is little evidence of an association between drug use and occupational achievement. Unlike Kandel *et al.* (1995), we find no significant negative associations between past or current drug use (soft or hard) and achievement - if anything, the lagged effect is positive. A mild positive association is also found between past soft drug use and occupational achievement for the younger cohort.

We summarise these results in figures 1-4, where solid arrows reflect a significant current effect and dashed arrows represent significant effects of past usage.

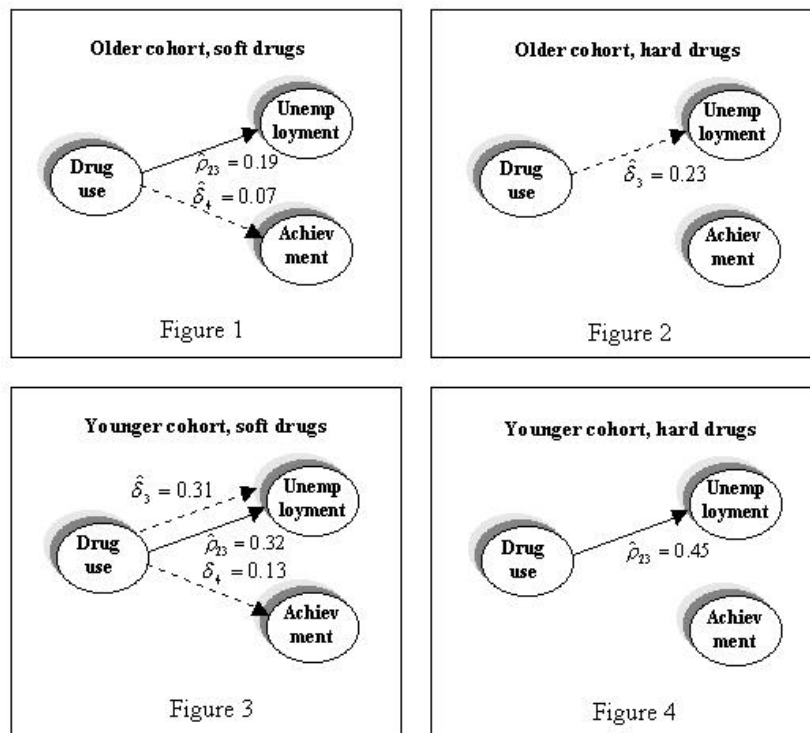


Figure 1:



It is difficult to provide a conclusive interpretation of these results. The positive association between past soft drug use and current labour market achievement may simply reflect unobserved attributes (such as personality) that are correlated with both past drug use and productivity (Register and Williams 1992, Kaestner 1994a). Alternatively, it may be the observable counterpart of a normal demand relationship, with past drug consumption positively related to past income, which is proxied by current occupational achievement. However, the strong evidence of a persistent adverse effect of (particularly hard) drug use on employment prospects is striking, so it is reasonable to conclude that our policy concern should be focused mainly on the unemployment effects of drug use, rather than its effect on the productivity of those who are in work. The fact that both soft and hard drugs appear to be associated with an increased probability of unemployment suggests that it might be rash to pursue any policy reform (such as decriminalisation) which is likely to lead to an increased use of soft drugs .

## References

- [1] Becker, G. and Murphy, K. 1988 A theory of rational addiction. *Journal of Political Economy*, vol. 96, no. 4, pp. 675-700
- [2] Burgess, S. M., and Propper, C. 1998. Early Health Related Behaviours and Their Impact on later Life Chances: Evidence from the US. *Centre for Analysis of Social Exclusion, CASE paper 6*, London School of Economics
- [3] Culyer, A. 1973. Should Social Policy Concern Itself with Drug Abuse? *Public Finance Quarterly*, vol. 1, no. 4, pp. 449-456
- [4] Farrell, M., Strang, J., Neeleman, J., and Reuter, P. 1994. Policy on Drug Misuse in Europe. *The British Medical Journal*, vol. 308, no. 6929, pp. 609-610
- [5] Gill, A., and Michaels, R. 1991. The Determinants of Illegal Drug Use. *Contemporary Policy Issues*, vol. 9, no. 3, pp. 93-105
- [6] Gill, A., and Michaels, R. 1992. Does Drug use Lower Wages. *Industrial and Labor Relations Review*, vol. 45, no. 3, pp. 419-434
- [7] Johnston, L., O Malley, P. and Bachman, J. 1988. *Illicit Drug Use, Smoking and Drinking by America s High School Students, and Young Adults*. Washington: US Department of Health and Human Services

- [8] Kaestner, R. 1991. The Effects of Illicit Drug Use on the Wages of Young Adults, *Journal of Labor Economics*, vol. 9, no. 4, pp.381-412
- [9] Kaestner, R. 1994a. The Effect of Illicit Drug Use on the Labour Supply of Young Adults, *Journal of Human Resources*, vol. 29, no. 1, pp. 126-155
- [10] Kaestner, R. 1994b. New Estimates of the Effects of Marijuana and Cocaine Use on Wages, *Industrial-and-Labor-Relations Review*, vol. 47, no. 3, pp.454-470
- [11] Kandel, D. 1980. Drug and Drinking Behaviour Among Youth, *Annual Review of Sociology*, vol. 6, pp 235-283
- [12] Kandel, D. 1984. Marijuana Users in Young Adulthood. *Archives of General Psychiatry*, vol. 41, pp 200-209
- [13] Kandel, D., Chen, K., and Gill, A. 1995. The Impact of Drug Use on Earnings: A Life-Span Perspective. *Social Forces*, vol. 74, no. 1, pp243-270
- [14] Labouvie, E. 1996. Maturing Out of Substance Abuse: Selection and Self-Correction, *Journal of Drug Issues*, vol. 6, no. 2, pp. 457-476
- [15] MacDonald, Z. 1997. The Demand for Illicit Drugs in the UK: Survey Evidence. *Discussion Paper in Public Sector Economics, 97/5*, University of Leicester
- [16] Ramsay, M., and Percy, A. 1996. Drug Misuse Declared: Results from the 1994 British Crime Survey . *Home Office Research Study 151*, London: Home Office Research and Statistic Department
- [17] Ramsay, M., and Percy, A. 1997. A National Household Survey of Drug Misuse in Britain: a Decade of Development. *Addiction*, vol. 92, no. 8, pp. 931-937
- [18] Ramsay, M., and Spiller, J. 1998. Drug Misuse Declared in 1996: Key Results from the British Crime Survey . *Home Office Research Study 172*, London: Home Office Research and Statistic Department
- [19] Register, C., and Williams, D. 1992. Labor Market Effects of Marijuana and Cocaine Use among Young Men. *Industrial and Labor Relations Review*, vol. 45, no. 3, pp. 435-448

- [20] Sickles R & Taubman P, 1991, *Who Uses Illegal Drugs?*, American Economic Review, vol. 81, no. 2, pp. 248-251
- [21] Tregoning, D. 1998. Home Office Addicts Index No Longer Exists. *The British Medical Journal*, vol. 316, no. 7125, p. 151
- [22] White, A., and Malbon, G. 1995. *1994 British Crime Survey: Technical Report*. London: OPCS Social Survey Division

## Appendix 1 Sample properties

Table A1: Drug Categories in the BCS

Drug	Class	Also known as:
Cocaine	A	Coke , Charlie
Crack	A	Rock
Ecstasy/MDMA	A	E , Eve , Dove
Heroin	A	Smack , H , Scag ,
LSD	A	Acid
Magic Mushrooms	A	Liberty Cap , Mushies
Methadone/Physeptone	A	Meth , Phy
Amphetamines	B	Speed , Phet , Whizz , Sulph
Cannabis	B	Pot , Blow , Dope , Grass , Hash , Draw , Black , Ganja , Marijuana , Herb
Tranquillisers/ Temapzepam	C	Jellies , Eggs , Rugby Balls
Amyl Nitrate	n/a	Poppers , Rush
Anabolic Steroids	n/a	Roids
Glues/Solvents		
Unknown Pills/Powders		
Unknown substance smoked		
Substance thought to be a drug		

Table A2: Variable and descriptive statistics

Variable	Description	Mean	Std. Dev
AGE	Chronological age	37.123	10.762
AEVER	1 = has used class A drug ever	0.090	0.286
BEVER	1 = has used class B/C drug ever	0.249	0.433
APSTONLY	1 = has used class A drug in past only	0.073	0.261
BPSTONLY	1 = has used class B/C drug in past only	0.165	0.371
ARECENT	1 = has used class A drug in past year	0.016	0.127
BRECENT	1 = has used class B/C drug in past year	0.085	0.279
ASIAN	1 = Asian (Indian, Pakistani, Bangladeshi)	0.063	0.242
BLACK	1 = Black (African, Caribbean, other)	0.070	0.255
OTHER	1 = Other, Chinese, none	0.021	0.144
WHITE	1 = white	0.847	0.360
DEGREE	1 = degree or higher	0.160	0.367
NON-DEGREE	1 = has other quali cation	0.648	0.478
EDU1	1 = degree or higher	0.160	0.367
EDU2	1 = Teaching/nursing quali cation, HND, BTEC	0.140	0.347
EDU3	1 = A levels, ONC, C&G advance.	0.115	0.319
EDU4	1 = high grade GCE/GCSE ,CSE grade 1, craft	0.257	0.437
EDU5	1 = low grade GCE/GCSE/CSE, clerical	0.086	0.280
EDU6	1 = other quali cation	0.051	0.220
EDU7	1 = no formal quali cations	0.192	0.394
FAMTYP1	1 = single adult household	0.165	0.371
FAMTYP2	1 = 2 adult household	0.274	0.446
FAMTYP3	1 = 3 or more adult household	0.139	0.346
FAMTYP4	1 = lone parent household	0.047	0.212
FAMTYP5	1 = two adults plus 1 or 2 children	0.272	0.445
FAMTYP6	1 = other household type	0.120	0.325
INCITY	1 = lives in inner city area	0.430	0.490
CHURCH	1 = belongs to church/synagogue/mosque/etc.	0.223	0.426
MALE	1 = male	0.534	0.499
MARRIED	1 = married	0.558	0.497
SNGLFMAL	1 = single female	0.224	0.417
SNGLMALE	1 = single male	0.218	0.413
MARRFMAL	1 = married female	0.242	0.428
MARRMALE	1 = married male	0.316	0.465
UNEMPL	1 = currently unemployed	0.070	0.255
SOCLASS	social class category (0 to 4)	2.927	1.136

## Appendix 2 Probabilities of observational outcomes

Under the model specification (10)-(17), the 15 possible observational outcomes occur with the following probabilities.

(i) *Never a drug user, currently unemployed*

$$\Pr(d_1^* \leq 0, d_2^* \leq 0, u^* > 0) = [1 - \Phi(x_1\beta_1)] \Phi(-x_2\beta_2, x_3\beta_3; -\rho_{23})$$

(ii) *Never a drug user, currently employed with occupational status  $j$  ( $j = 1..4$ )*

$$\begin{aligned} \Pr(d_1^* \leq 0, d_2^* \leq 0, u^* \leq 0, C_{j-1} \leq a^* < C_j) = \\ [1 - \Phi(x_1\beta_1)] [\Phi(-x_2\beta_2, -x_3\beta_3, C_j - x_4\beta_4; \rho_{23}, \rho_{24}, \rho_{34}) \\ - \Phi(x_2\beta_2, -x_3\beta_3, C_{j-1} - x_4\beta_4; \rho_{23}, \rho_{24}, \rho_{34})] \end{aligned}$$

(iii) *Early but not current drug user, currently unemployed*

$$\Pr(d_1^* > 0, d_2^* \leq 0, u^* > 0) = \Phi(x_1\beta_1) \Phi(-x_2\beta_2 - \delta_2, x_3\beta_3 + \delta_3; \rho_{23})$$

(iv) *Early but not current drug user, currently employed with occupational status  $j$  ( $j = 1..4$ )*

$$\begin{aligned} \Pr(d_1^* > 0, d_2^* \leq 0, u^* \leq 0, C_{j-1} \leq a^* < C_j) = \\ \Phi(x_1\beta_1) [\Phi(-x_2\beta_2 - \delta_2, -x_3\beta_3 - \delta_3, C_j - x_4\beta_4 - \delta_4; \rho_{23}, \rho_{24}, \rho_{34}) \\ - \Phi(-x_2\beta_2 - \delta_2, -x_3\beta_3 - \delta_3, C_{j-1} - x_4\beta_4 - \delta_4; \rho_{23}, \rho_{24}, \rho_{34})] \end{aligned}$$

(v) *Current drug user, currently unemployed*

$$\begin{aligned} \Pr(d_1^* > 0, u^* > 0) + \Pr(d_1^* \leq 0, d_2^* > 0, u^* > 0) = \Phi(x_1\beta_1) \Phi(x_3\beta_3 + \delta_3) \\ + [1 - \Phi(x_1\beta_1)] \Phi(x_2\beta_2 + \delta_2, x_3\beta_3 + \delta_3; \rho_{23}) \end{aligned}$$

(vi) *Current drug user, currently employed with occupational status  $j$  ( $j = 1..4$ )*

$$\begin{aligned}
& \Pr(\max(d_1^*, d_2^*) > 0, u^* \leq 0, C_{j-1} \leq a^* < C_j) \\
& \quad = \Pr(d_1^* > 0, u^* \leq 0, C_{j-1} \leq a^* < C_j) \\
& \quad + \Pr(d_1^* \leq 0, d_2^* > 0, u^* \leq 0, C_{j-1} \leq a^* < C_j) \\
& = \Phi(x_1\beta_1) [\Phi(-x_3\beta_3 - \delta_3, C_j - x_4\beta_4 - \delta_4; \rho_{34}) \\
& \quad - \Phi(-x_3\beta_3 - \delta_3, C_{j-1} - x_4\beta_4 - \delta_4; \rho_{34})] \\
& + [1 - \Phi(x_1\beta_1)] \times \\
& \quad [\Phi(x_2\beta_2 + \delta_2, -x_3\beta_3 - \delta_3, C_j - x_4\beta_4 - \delta_4; -\rho_{23}, -\rho_{24}, \rho_{34}) \\
& \quad - \Phi(x_2\beta_2 + \delta_2, -x_3\beta_3 - \delta_3, C_{j-1} - x_4\beta_4 - \delta_4; -\rho_{23}, -\rho_{24}, \rho_{34})]
\end{aligned}$$

In these expressions,  $\Phi$  is the cdf of either the univariate, bivariate or trivariate standard normal distribution. In the last two cases, correlation parameters are also specified. The notation here is  $\rho_{ij} = \text{corr}(\varepsilon_i \varepsilon_j)$ .

### Appendix 3 Estimation results

Table A3i

The probability of past drug use: estimates for 16-19 cohort

Coefficient ( $\beta_1$ )	Class A	S.E.	Class B/C	S.E.
MALE	0.377***	0.084	0.302***	0.082
AGE	-0.071	0.118	0.280**	0.168
DEGREE	-0.369**	0.174	0.138	0.170
NON-DEGREE	-0.154	0.129	-0.045	0.133
BLACK	-0.011	0.144	-0.075	0.145
ASIAN	-0.692***	0.232	-0.533***	0.186
OTHER	0.181	0.264	0.091	0.252
CHURCH	-0.114	0.124	-0.314***	0.114
CONSTANT	-1.066***	0.317	-1.457***	0.505
n	1906		1906	

Table A3ii

The probability of current drug use: estimates for 16-19 cohort

Coefficient ( $\beta_2$ )	Class A	S.E.	Class B/C	S.E.
MALE	0.312***	0.128	0.340***	0.090
AGE	-0.689***	0.194	-0.632***	0.148
INCITY	0.064	0.118	0.085	0.074
DEGREE	0.195	0.257	0.287**	0.167
NON-DEGREE	-0.062	0.190	0.066	0.133
BLACK	-0.052	0.219	-0.005	0.133
ASIAN	-0.360*	0.263	-0.846***	0.220
OTHER	0.636**	0.308	0.235	0.226
FAMTYPE1	0.108	0.225	0.237	0.158
FAMTYPE2	-0.026	0.199	0.017	0.148
FAMTYPE3	-0.016	0.199	0.037	0.145
FAMTYPE4	-0.078	0.422	0.418**	0.215
FAMTYPE5	-0.032	0.218	0.081	0.151
MARRIED	-0.400**	0.208	-0.517***	0.123
CHURCH	-0.401**	0.238	-0.033	0.128
CONSTANT	-0.085	0.458	0.158	0.296
past use ( $\hat{\delta}_2$ )	-3.056	282.904	0.718*	0.526
n	1906		1906	

Note: \*\*\* = significant at 1% level, \*\* = significant at 5% level, \* = significant at 10% level



Table A3iii

The probability of current unemployment: estimates for 16-19 cohort

Coefficient ( $\beta_3$ )	Class A	S.E.	Class B/C	S.E.
MALE	0.913***	0.158	0.872***	0.153
AGE	0.072	0.163	0.043	0.161
INCITY	0.246***	0.107	0.260***	0.111
EDU1	-0.787***	0.207	-0.804***	0.206
EDU2	-0.885***	0.221	-0.874***	0.230
EDU3	-0.666***	0.195	-0.666***	0.192
EDU4	-0.537***	0.147	-0.521***	0.149
EDU5	-0.399***	0.183	-0.430***	0.183
EDU6	-0.379**	0.233	-0.312*	0.232
BLACK	0.512***	0.143	0.529***	0.143
ASIAN	0.181	0.192	0.242	0.194
OTHER	-0.094	0.384	-0.029	0.350
FAMTYPE1	0.899***	0.226	0.887***	0.225
FAMTYPE2	-0.116	0.231	-0.114	0.227
FAMTYPE3	-0.723***	0.328	-0.726***	0.329
FAMTYPE4	1.432***	0.256	1.429***	0.251
FAMTYPE5	0.320**	0.184	0.298**	0.185
MARRIED	-0.038	0.157	-0.035	0.156
CONSTANT	-2.166***	0.463	-2.164	0.466
past use ( $\hat{\delta}_3$ )	0.019	0.152	0.308**	0.163
n	1906		1906	

Note: \*\*\* = significant at 1% level, \*\* = significant at 5% level, \* = significant at 10% level

Table A3iv

The probability of occupational achievement: estimates for 16-19 cohort

Coefficient ( $\beta_4$ )	Class A	S.E.	Class B/C	S.E.
AGE	0.269***	0.095	0.261***	0.095
INCITY	-0.089*	0.058	-0.092**	0.058
EDU1	2.122***	0.135	2.108***	0.134
EDU2	1.420***	0.127	1.417***	0.127
EDU3	1.008***	0.123	1.006***	0.123
EDU4	0.713***	0.105	0.709***	0.105
EDU5	0.522***	0.124	0.509***	0.125
EDU6	0.492***	0.160	0.505***	0.157
BLACK	-0.297***	0.114	-0.294***	0.115
ASIAN	0.351***	0.109	0.365***	0.111
OTHER	-0.031	0.187	-0.037	0.183
FAMTYPE1	0.421***	0.149	0.420***	0.144
FAMTYPE2	0.375***	0.097	0.375***	0.097
FAMTYPE3	0.248***	0.102	0.252***	0.103
FAMTYPE4	-0.325*	0.206	-0.329**	0.200
FAMTYPE5	0.169**	0.101	0.171**	0.101
SNGLMALE	-0.148*	0.099	-0.139	0.099
SNGLFMAL	0.035	0.112	0.052	0.110
MARRFMAL	-0.015	0.111	0.006	0.110
past use ( $\hat{\delta}_4$ )	0.037	0.083	0.132**	0.079
C1	-0.168	0.279	-0.152	0.277
C2	0.849***	0.276	0.866***	0.275
C3	2.513***	0.280	2.531***	0.280
n	1906		1906	

Note: \*\*\* = significant at 1% level, \*\* = significant at 5% level, \* = significant at 10% level