

Immunity and Susceptibility in Illicit Drug Initiation in Israel

Michael Beenstock^{1,3} and Giora Rahav²

We model the initiation process into cannabis and hard drugs using long-term survivor analysis. This approach hypothesizes two sub-populations: a population that is "immune" to drugs, who will never use them no matter how long they live, and a population which is "susceptible" to drugs for whom it is a matter of time until they begin to use drugs. We use data for Israel to test competing hypotheses concerning the timing of drug use initiation and the determination of susceptibility. Cigarettes do not significantly affect immunity to drugs, but they tend to speed up the rate of initiation for those who smoke them. This implies that anti-smoking policy may only delay drug use initiation without affecting immunity. Finally, drug initiation in Israel is not explained by variables that are traditionally associated with criminality.

KEY WORDS: illicit drugs; gateway theory; long-term survivor analysis.

1. INTRODUCTION

Drug use is often assumed to express an underlying problem that manifests itself in the use of psychoactive substances including cigarettes, alcohol and illicit drugs. The same problem is also assumed to express itself in various aspects of substance use, including age at initiation, and intensity and duration of use. According to this view, the factors that increase drug use are also expected to lower the age of initiation into drugs, and raise the age of termination. However, Knupfer and Room

¹Department of Economics, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel.

²Department of Sociology, Tel Aviv University.

³To whom correspondence should be addressed.

E-mail: msbin@mscc.huji.ac.il

(1967) have shown that members of ethnic or religious groups who tend to avoid alcohol are more likely to engage in excessive drinking if they drink. This serves to remind us that the processes of initiation, use and termination may differ.

Indeed, the need to study drug use initiation as a phenomenon in its own right is suggested by the growing literature on drug use “careers”, which claims that the determinants of behavior are likely to vary over the life-cycle of drug use. Fergusson and Horwood (1997) investigated the effects of early initiation into cannabis, but did not model the initiation process. Since early initiation is a negative prognosticator of future drug use, their findings emphasize the importance of acquiring a deeper understanding into the initiation process itself. In this paper, we therefore focus exclusively upon the initiation process, and discuss theoretical and methodological problems that are associated with it.

Previous studies of the initiation process of illicit drug use are scarce. Segal (1991), Chilcoat and Schutz (1996), and Johnson and Gerstein (1998) have investigated the hazard functions in the United States for different illicit substances by gender and cohort. The initiation hazard typically rises sharply between the ages of 13 and 15, peaks in the late teens before declining rapidly to zero by age 25. Evidence on cohort effects seems to be mixed with some studies claiming that later birth cohorts have lower hazards, while others claim the opposite.

Probably the most important investigations of illicit drug initiation are Kandel and Logan (1984) and Yamaguchi and Kandel (1984), who estimated multivariate survival and hazard functions. They found that gender, age of alcohol and cigarette initiation and a series of environmental and interpersonal characteristics (including mother’s use of drugs and attitudes to drugs) were significant predictors of age at drug use initiation. We think that their approach poses a number of problems. First, they defined initiation in terms of life-time use in excess of 10 times, which in our view focuses on the transition from experimental use to occasional use rather than from non-use to use. Secondly, the standard survival model that they used assumes that drug initiation, like death, is inevitable and a matter of time. If at the time of the survey an individual reports that he has never used drugs, i.e. the observation is censored, the standard model automatically assumes that it is simply a matter of time until initiation occurs. It is unreasonable to assume *ex hypothesi* that drug initiation is eventually inevitable (unless death chances to intercede). This is particularly true of drug use where survival distributions, such as those presented in Figs. 1–3, typically bottom out at a proportion that is (thankfully) well in excess of zero. Common sense suggests that this arises not because people do not live long enough, but because the population

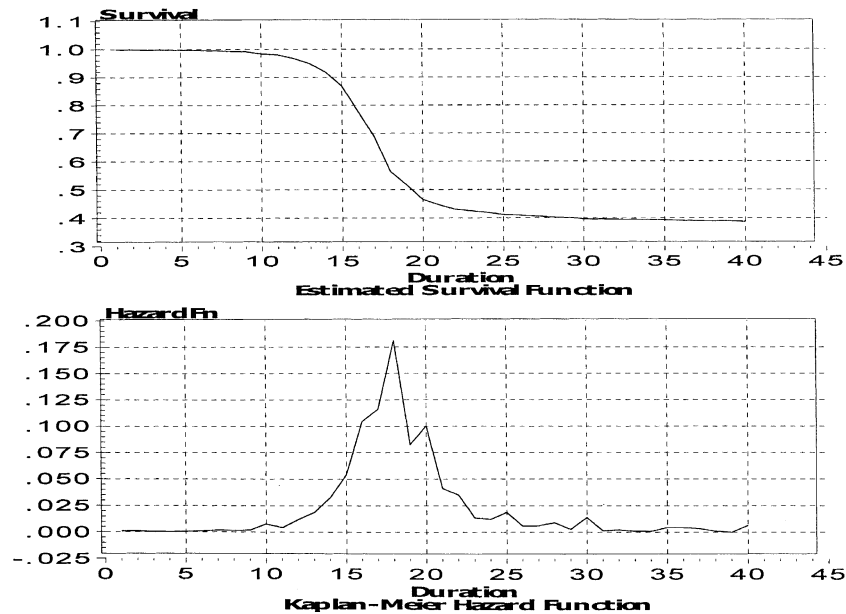


Fig. 1. Cigarettes.

comprises individuals who would not initiate drug use no matter how long they lived.⁴

We propose a more general survival model, which splits the population into two groups and recognizes that drug initiation is not inevitable, because there are individuals who would never use drugs no matter how long they live. In this model, censored observations come from two populations; an “immune population” that would never use drugs no matter how long they lived, and a “susceptible population” for whom drug initiation is a matter of time. In the bio-statistical literature these split population or mixture models are known as “long-term survivor models” (LTSM⁵).

A third criticism of Yamaguchi and Kandel (1984) concerns the parametric assumptions (exponential) of their model. The exponential

⁴It might be argued that Yamaguchi and Kandel (1984), by using age as a control variable, mitigate this problem. We argue that controlling for age captures a cohort effect since older individuals are born earlier; it does not solve the problem that we raise.

⁵Maller and Zhou (1996) record that LTSM dates back to 1949 and that applications to criminal behavior date back to 1969. Subsequently, there have been numerous applications of LTSM in the fields of medicine, engineering, biology, criminology, and economics. Apparently, the application of LTSM to the study of illicit drug use is new. However, Douglas and Hariharan (1994) applied LTSM to cigarette initiation.

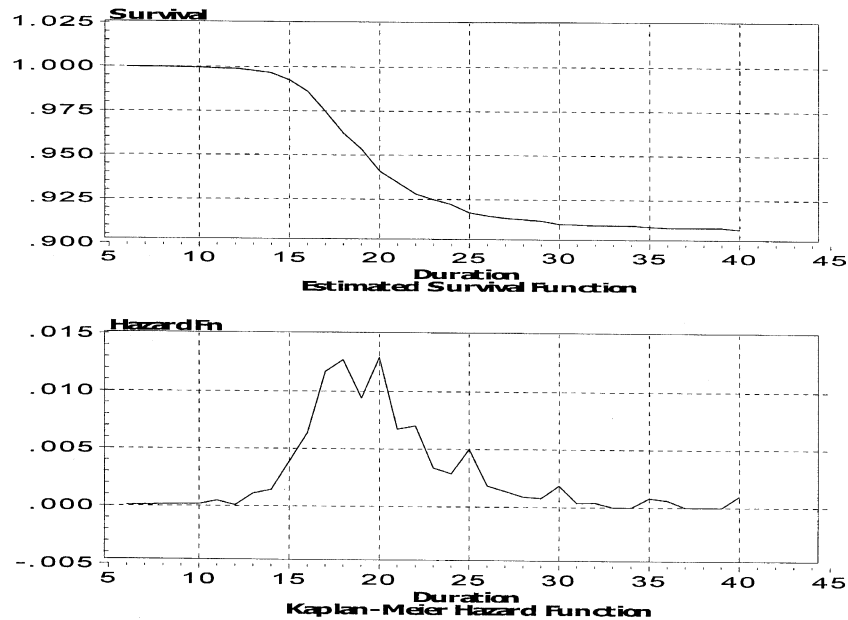


Fig. 2. Cannabis.

distribution assumes that the hazard of initiating drugs is constant; the probability of initiating is the same at all ages. In Section 3, we show that in Israel, as elsewhere, the hazard is variable. It is well known (Heckman and Singer, 1984) that in contrast to ordinary least squares regressions, failure time regressions of this variety are particularly sensitive to the parametric assumptions that are used. The unwitting researcher, who applies some inappropriate survival distribution, might reach the conclusion that a given covariate has a positive effect upon the outcome of interest when the opposite happens to be true. We therefore suggest that in the interest of robustness the drug initiation process be tested using alternative statistical models and parametric assumptions.

We begin by investigating the process of cannabis initiation. Thereafter we investigate hard drug initiation (first use of cocaine, heroin, LSD and amphetamines). Apart from investigating the effects of individual characteristics (such as ethnic background, gender and birth cohort) on drug initiation and immunity to drugs, we focus upon “Gateway” effects. Is it the case, for example, that for given characteristics, early initiation of cigarettes is associated with early initiation of cannabis? Is it also the case that early initiation of cannabis is associated with early initiation of hard drugs?

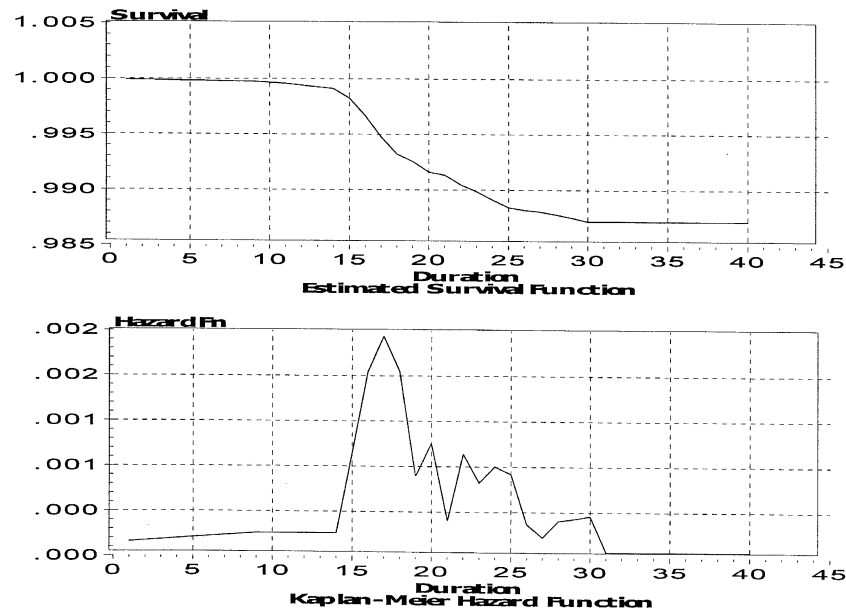


Fig. 3. Hard drugs.

2. HYPOTHESES

The available literature provides few guidelines for forming hypotheses about the timing of drug initiation. The major ones are that initiation is concentrated among adolescents and young adults, and that the most common order of initiation, as suggested by Gateway Theory, is legal substances⁶ followed by cannabis and then other drugs. The literature provides even fewer guidelines regarding the factors that determine immunity and susceptibility. Indeed, the differences between immunity, susceptibility and initiation in illicit drug use have been blurred and even confused. One of our current objectives is to distinguish between them, and to raise them as legitimate and important topics in the study of addictions.

If initiation is largely a stage of experimentation rather than sustained use, we can hypothesize that early initiators will be predominantly male adolescents, a population that is well known for sensation and novelty seeking. Since religion is an important system of social control we should expect that the probability of early initiation varies inversely with religious observance and devoutness.

⁶Israel is unusual in that alcohol does not serve as a legal gateway substance. However, cigarettes serve as a gateway substance (Beenstock and Rahav, 2002).

An alternative hypothesis sees illicit drug use as a form of criminal or deviant behavior. According to this hypothesis, the factors predicting criminal behavior are also likely to predict early initiation of drugs. Since deviant behavior, from psychiatric problems to crime, is more prevalent among lower socio-economic groups (Dohrenwend *et al.*, 1992), it is reasonable to expect that these groups will be particularly associated with earlier onset of drug use. Among Israeli Jews, this would suggest earlier drug use among individuals with less than high school education.

Furthermore, as there is a clear association in Israel (Rahav, 1998) between country of origin and crime and deviance, we may expect that drug initiation be related to country of origin.

If cannabis is viewed as a recreational activity, then like most other forms of recreation it should characterize middle and upper class individuals. This suggests that early initiation should be greater among educated Ashkenazi Jews, especially from Western Europe and the Americas. Finally, Gateway Theory suggests that susceptibility and initiation in the case of cannabis should be related to previous consumption of cigarettes, and in the case of hard drugs they should be related to previous consumption of cannabis. In Sections 5 and 6 we investigate these hypotheses using the methodologies discussed in Section 3 and the data described in Section 4.

3. METHODOLOGY

In this Section we discuss how LTSM permit the empirical distinction between immune and susceptible populations.⁷ LTSMs postulate that censored individuals (i.e. who at the time of data collection had not initiated drugs) come from two populations; those that are susceptible to drugs, and those that are immune to drugs. LTSM is inherently a mixture model because it mixes two distributions: the probability distribution for immunity and the survival distribution for the susceptible.

3.1. Long-Term Survivor Models

Let t_i denote the time at which drug use is initiated by individual i . The survival function outside of drugs is defined as $S(t) = 1 - F(t)$, and the hazard function for drug initiation is $h(t) = f(t)/S(t)$. In the data some of the observations will be censored, i.e. at the time the data are gathered the event of interest, drug initiation, has not yet occurred. The standard estimation procedure defines an indicator δ_i that takes a value of 1 if the observation i is not censored, and 0 otherwise. The log-likelihood function that results is:

⁷Maller and Zhou (1996) provide a comprehensive review of LTSM.

$$\log L = \sum_{\delta=1} \log[f(t_i/\theta)] + \sum_{\delta=0} [S(t_i/\theta)] \quad (1)$$

where θ is a vector of parameters to be estimated. Eq. (1) conventionally implies that it is a matter of time before the event of interest happens to the censored individuals, i.e. nobody is immune to drugs, unless by chance death happens to intercede.

Long-term survivor analysis defines P_i to be the probability that individual i is not a long-term survivor, i.e. for him it is a matter of time until the event of interest happens; he is susceptible to the event. However, with probability $1 - P_i$ he is immune to the event in the sense that the event will not occur in his lifetime no matter how long he lives. The non-censored observations in Eq. (1) contribute to the identification of $f(\cdot)$ because, for them we know that they are not long-term survivors, i.e. they are susceptible. For the censored observations, there are two possibilities; they belong to the immune population with probability $1 - P_i$, and with probability P_i the event will eventually occur after the date of censoring because they belong to the susceptible population.

The data may be used to estimate the underlying survival density, $f(t/\theta)$, for the susceptible population as in standard survival analysis, as well as the immunity/susceptibility model for $P(\phi)$, where ϕ is a vector of parameters of the immunity model. The log-likelihood function for the mixture model that results is:

$$\log L = \sum_{\delta=1} \log[P_i(\phi)f(t_i/\theta)] + \sum_{\delta=0} \log[1 - P_i(\phi) + P_i(\phi)S(t_i/\theta)] \quad (2)$$

Eq. (2) is richer than Eq. (1) in that it views the data as coming from a mixture of two distributions, $P(\cdot)$ and $f(\cdot)$, instead of $f(\cdot)$ alone.

3.2. Specification of the Model

It remains to specify the immunity model for P , and the initiation density $f(\cdot)$ in Eq. (2). We begin with the latter. Since the density, the hazard and survival functions are related we discuss the specification of $h(\cdot)$ rather than $f(\cdot)$. Survival analysis comes in three main forms, parametric, semi-parametric, and non-parametric. The latter does not have sufficient structure for our present purposes. The former specifies a loglinear survival distribution with parameters λ and p . The Weibull density implies that the hazard function is $h = \lambda p(\lambda t)^{p-1}$, which states that the hazard increases monotonically when $p > 1$ (positive duration dependence) or it decreases (negative duration dependence) when $p < 1$. A special case of the Weibull density is the exponential, which arises when $p = 1$ and implies a constant

hazard rate equal to λ . The lognormal density, $f = (p/t)\phi[\text{pln}(\lambda t)]$, and the log-logistic density may give rise to non-monotonic hazard functions. For example, in the log-logistic case $h = \lambda p(\lambda t)^{p-1}/[1 + (\lambda t)^p]$, i.e. the hazard function may be U-shaped or \cap -shaped.

In the semi-parametric case we follow Cox's model of proportional hazards by assuming that the hazard function is $h = \lambda h_0(t)$, where h_0 denotes the baseline hazard. Since the latter is determined empirically by the data the Cox model runs less of a risk of specifying some inappropriate hazard function. This will be important if the true hazard function is none of the parametric functions mentioned in the previous paragraph, or any other parametric function. An obvious disadvantage of the Cox model is that it assumes that h_0 and λ are independent, i.e. that the hazards are proportional to λ . By contrast parametric survival analysis does not assume that λ and p are independent in the data.

We assume that $\lambda_i = \exp(-X_i\beta)$ where X is a vector of covariates hypothesized to determine the hazard for individual i . The coefficient vector β may be estimated by maximum likelihood along with p in the parametric case and h_0 in the semi-parametric case.

The distribution of susceptibility to drug use in the population is, of course, unknown. Here we assume two cases. In the first, susceptibility is assumed to have a logistic distribution, while in the second it is assumed to be normally distributed. The former has more extreme values than the latter, which may be more appropriate for modeling deviant behavior. The susceptibility model for P may therefore be specified either as a logit model in which case:

$$P_i = \frac{1}{1 + \exp(\alpha Z_i)} \quad (3)$$

where Z^8 is a vector of variables hypothesized to determine susceptibility to drugs, and α is the associated coefficient vector to be estimated. Or, if susceptibility is assumed to be normally distributed, it may be specified as a probit model in which case $P_i = 1 - \Phi(\alpha Z_i)$, where $\Phi(\cdot)$ is the cumulative density function for the standard normal. Note that this normalization implies that susceptibility varies inversely with Z if $\alpha > 0$. Note also that the variables hypothesized to determine the timing of the event (X), are different from those hypothesized to affect immunity (Z). However, even if $Z = X$ all the parameters of the model are identified parametrically because f and P have different distributions.

Larson and Dinse (1985) proposed an alternative mixture model in which the logarithm of the baseline or null hazard is assumed to be

⁸Schmidt and Witte (1989) and Douglas and Hariharan (1994) among others assume that the X and Z variables are the same. We allow them to differ.

piecewise-linear over failure time sub-intervals instead of fully parametric (e.g. Weibull etc), and in which the competing risks are assumed to have a logistical distribution. Subsequently, Kuk and Chen (1992) proposed a mixture model in which the baseline has Cox's proportional hazards representation.⁹ The advantage of this semi-parametric approach is that it does not make what may be arbitrary assumptions about the nature of the hazard function.¹⁰ We experiment with this semi-parametric mixture methodology in Section 5 in addition to the fully parametric approach.

3.3. Misspecification Tests

As mentioned in Section 1, the very fact that survival functions for drug use tend to bottom out well above zero testifies to the existence of an immune population. Maller and Zhou (1996, cap 2) have suggested formal tests for the presence of immunes and sufficient follow-up. Schmidt and Witte (1989) test for immunes by comparing the estimated likelihoods of LTSMs and conventional failure time models in which everyone is assumed to be susceptible. In our drug use data the evidence in favor of an immune population turns out to be overwhelming.

Maller and Zhou (1996, cap 5) have also suggested an order statistic for testing goodness-of-fit. This test is based on the correlation (r) between the order of failure predicted by the estimated model and the order of failure from the Kaplan–Meier (KM) estimator. The greater the correlation, the better is the goodness-of-fit. The correlation (r) is expected to be higher the larger is the sample size, the greater is the rate of susceptibility, and the lower is the rate of censoring. This test uses censored as well as uncensored data. We apply this test below.¹¹

When the data are binary it is well known that the average predicted probability from an estimated logit model must be equal to the proportion of the event in the data provided the model includes an intercept term.¹² Therefore, if susceptibility is modeled as in Eq. (3) it may be natural to expect that the average predicted rate of susceptibility should be similar to the KM asymptote. For example, Fig. 2 below shows that the KM estimate of the asymptote for cannabis susceptibility is about 9%. It may be shown

⁹The Larson–Dinse model is equivalent to the Cox model when the subintervals become as frequent as the number of failure times in the data.

¹⁰Kuk and Chen (1992) obtained quite different estimates of immunity rates from the proportional hazard and Weibull specifications.

¹¹Escarela *et al.* (2000) do not apply this test. Instead, they validate the results by randomly splitting the sample into two, which is similar in spirit to a Chow test.

¹²Note that this result has not been proved for the probit model.

that in LTSMs there is no guarantee that the average predicted probability of susceptibility will be 9%. It may be higher or lower. This arises from the fact that the data are not binary. Indeed, in many of the examples in Maller and Zhou (1996) the predicted rate of susceptibility is often quite different from its KM counterpart.¹³ Note that the order statistic of a model might be satisfactory even if it over or underpredicts the rate of susceptibility. This happens because the predicted order of failure may be satisfactory even if the predicted probability of failure happens to be biased.

In addition to Maller and Zhou's r statistic for goodness-of-fit we also compare the predicted survival curve to the actual survival curve. This comparison reveals whether an over or under-estimate of the rate of susceptibility results from a poor model fit over the range of the data, or whether it results from adverse extrapolation beyond the data. The former is obviously more serious than the latter. For example, in our data the maximum age is about 40 years. We ask whether the estimated LTSM tracks the survival curve up to aged 40, even if extrapolating beyond aged 40 provides what seems to be a biased estimate of the rate of susceptibility.

While the choice between LTSM and the null is quite straightforward, matters are more complicated when we compare rival specifications of the hazard function. If the KM hazard function is non-monotonic, as is the case in our data, a Weibull specification is unlikely to fit the data well because it restricts the hazard to be either increasing or decreasing. The natural candidates here are the log-logistic or log-normal distributions or proportional hazards which do not restrict the hazard function to be monotonic.

3.4. Susceptible Hazard Analysis

The total population comprises the immune and the susceptible. Since P denotes the probability of susceptibility, the latter is equal to PS_0 and the former to $(1-P)S_0$. We distinguish between the drug initiation hazard for the population as a whole, i.e. $h = f/S$, and the initiation hazard for the susceptible population, i.e. $h^* = f/S^*$, where $S^* = S - (1-P)S_0$ denotes the number of survivors in the susceptible population. The two hazards are related as follows:

$$h^* = \frac{f}{S - (1 - P)S_0} = \frac{h}{1 - (1 - P) \frac{S_0}{S}} \quad (4)$$

¹³For example, on p150 in Maller and Zhou (1996) the predicted rate of susceptibility ranges between 0.43 and 0.53 when the KM rate is 0.42. Typically, the predicted rate of susceptibility is not reported, e.g. Schmidt and Witte (1989), Douglas and Hariharan (1994), and Escarela *et al.* (2000). A rare exception is Kuk and Chen (1992).

Eq. (4) states that the hazard rate for the immune population behaves quite differently to the hazard rate for the population as a whole, because the denominator contains S . For example, if $h = \lambda$ (the exponential case) $h^* = \lambda/[1-(1-P)S_0\exp(\lambda t)]$, which implies that the hazard rate increases for the susceptible despite the fact that it is constant for the population. Eq. (4) states that h^* will only be constant when $h = 1-(1-P)S_0/S$, which is a condition that will be rarely fulfilled. If, for example, the population hazard function is \cap -shaped, as it is in our data, the susceptible hazard function cannot be exponential. Alternatively, if $h^* = \lambda$, Eq. (4) states that the hazard function for the population is $h = \lambda e^{-\lambda t}$, which implies that the population hazard decreases despite the fact that it is constant for the susceptible. The choice of density function that should be used in Eq. (2) should be motivated by h^* rather than h . *Ex ante* we do not know who is immune and who is susceptible. *Ex post* estimates of LTSM provide estimates of the proportion of the population that is immune. We may then use these estimates to calculate h^* via Eq. (4).

3.5. Heterogeneity and Frailty

Conventional survival analysis assumes that the population is homogeneous and that the event of interest occurs randomly. Accelerated failure time analysis assumes that the heterogeneity is entirely observable and is captured by the covariates specified in the model. There is no theoretical reason why all the heterogeneity should be observable. The population may be heterogeneous (Heckman and Singer, 1984) in which different groups, whose characteristics are unobservable, have different degrees of susceptibility to the event of interest. Suppose, for example, that the population consists of two groups, the “frail” and the “strong”, that the event occurs more rapidly among the frail, and that the true distribution of the event is Weibull. Failure to recognize the existence of these two groups could bias the empirical analysis. The unwitting researcher, who mistakenly assumes homogeneity, may reach the incorrect conclusion that the distribution is not Weibull. Heckman and Singer go further and show that he may even reach mistaken conclusions about the direction of the effect of the covariates upon the hazard.

LTSM may be regarded as a form of heterogeneity in which the two groups are the immune and the susceptible. However, among the susceptible population there may be heterogeneity too. For example, there may be two groups, the “very susceptible” and the “less susceptible”. The appropriate LTSM may contain heterogeneity among the susceptible. By contrast, in Section 3.1 it was assumed that the susceptible population is homogeneous.

Just as the distinction between susceptibility and immunity deepens understanding of the initiation process, so may the distinction between the degree of susceptibility among the susceptible deepen this understanding. However, the former distinction is clearly the more important of the two.

We now consider the implications of heterogeneity for LTSM. Broadly speaking, two main approaches have emerged for modeling heterogeneity. The first makes parametric assumptions about the distribution of unobserved heterogeneity in the population. In our context, this means that the degree of susceptibility or frailty has a distribution, and that there is a continuity of groups. The second approach is non-parametric and is due to Heckman and Singer (1984), who assume that the distribution has a finite number of mass points, and that there is a finite number of groups. The number of such points may be thought of as the number of types or classes in the susceptible population. In our context, if there were two such mass points they would refer to the two classes mentioned above, the “more susceptible”, and the “less susceptible”.

Let v_i denote the degree of susceptibility of individual i , $S(t_i/v_i)$ denotes the survival function conditional on the degree of susceptibility, and $g(v)$ denotes the distribution of the degree of susceptibility among the susceptible. We illustrate the first approach by assuming for simplicity that the event of interest has a Weibull distribution in which case the hazard function for the susceptible is:

$$h^*(t) = \lambda p (\lambda t)^{p-1}$$

where λ and p are defined as in Section 3.2. The degree of susceptibility is assumed to have a gamma distribution:

$$g(v) = \frac{k^k}{\Gamma(k)} e^{-kv} v^{k-1}$$

with mean 1 and variance $v = 1/k$. It may be shown (Greene, 2000, pp. 946–947) that in this case the hazard function for the susceptible, after allowing for the effects of heterogeneity, is equal to:

$$h^\#(t) = \frac{h^*(t)}{1 + v(\lambda t)^p} \quad (5)$$

According to Eq. (5), $h^\#$ may be non-monotonic because t appears in both numerator and denominator. The numerator is of course monotonic because h^* is the hazard function for the Weibull. It is easy to see how the denominator acts to induce non-monotonicity because when $v = 1$ Eq. (5) is identical to the hazard function for the log–logistic distribution, which is

necessarily non-monotonic. If $v = 0$ there is no heterogeneity and Eq. (5) implies that $h^\# = h^*$, i.e. we revert to the Weibull hazard function, which is monotonic. This implies that the greater the degree of heterogeneity, i.e. the greater is v , the more likely it will be that $h^\#$ will be \cap -shaped.

This discussion shows that frailty among the susceptible may induce non-monotonic hazards of the type typically observed in drug initiation data. This is yet another reason for rejecting the Weibull distribution in the specification of the hazard function for the susceptible.

4. METHODS

The data come from three separate surveys carried out in Israel in 1989, 1992 and 1995 by the Israel Anti-Drug Authority (to whom we are grateful for making the data available). Since drug use is a rare event, we hope to learn more from three surveys than from one because this increases the number of drug users in the combined sample. It also enables us to investigate time trends and cohort effects in drug use because age and year of birth, which are obviously perfectly correlated in a given survey, are to some degree independent when we combine surveys from different years.¹⁴

Interviews were carried out face to face, but sensitive information on drug use was supplied on self-administered answer sheets. The population consists of Jews aged between 20 and 40. Details of the survey and summary statistics may be found in Rahav *et al.* (1996). Respondents were asked to provide information on consumption of cigarettes, alcohol and illicit drugs during the last week, last month, last year, and ever. They were also asked to report age of initial use of each substance. The total sample size over all three surveys is approximately 12,500.

Respondents provided basic demographic socio-economic data. Education was measured on a scale ranging from one (no formal education) to nine (higher academic degree). On this scale, the first three points correspond to elementary education, the second three to high school (from drop-outs to completers), and the final three to tertiary education. Ethnic origin (they are all Jews) is expressed by parents' country of origin, since Ashkenazi Jews come from Europe and the Americas, and Sephardic Jews from Asia and Africa. We also distinguish between natives and foreign born. Religious observance is measured on a subjective 5-point scale.

We represent age at initiation for each substance as reported by the respondents in the surveys. Thus, the initiation data are retrospective. Furthermore, we have no way of knowing how they defined initiation for

¹⁴Since undertaking the research two further surveys have been made available for the years 1998 and 2001.

themselves; for example, literally as first use or as continuous use. This kind of definition is quite common in the study of psychoactive substances and their consumption, e.g. Willens *et al.* (1997) and Tartar *et al.* (1995) in the case of disorder and dependence, Turnbull *et al.* (1990) and Luthar *et al.* (1992) in the case of psychiatric diagnosis, and De Wit *et al.* (1999) in the case of initiation. As mentioned, the definition differs substantively from the one used by Yamaguchi and Kandel (1984).

The data are subject to two possible types of error. Respondents may not remember accurately when they initiated their drug consumption. This error is likely to increase with the passage of time since initiation. Secondly, because drug use is sensitive respondents may bias their replies despite the fact that this information is provided anonymously. The response biases may be quite complex. Some may wish to deceive by recording that they initiated later than they did, while others might wish to brag by recording that they initiated earlier. We do not address these matters here. However, Tsibel (2000) has used our data to investigate them. She found that older people are more likely to misreport the age of drug initiation, as are people with more inconsistencies in their responses. She also found that heavy smokers (of cigarettes) are better at remembering when they initiated. Most importantly, she found that the parameter estimates obtained from models that assume that there is no data misreporting turn out to be very similar to the parameter estimates from models in which misreporting is assumed to take place. These conclusions suggest that even if misreporting occurs, it is unlikely to bias our findings.

Figures 1–3 plot the empirical (KM) survival and hazard functions for cigarettes, cannabis and hard drugs. These initiation distributions are estimated using the data in all three surveys, and indicate that some 60% eventually start smoking cigarettes, almost 9% eventually start using cannabis, and about 1.5% eventually use hard drugs. The tests suggested by Maller and Zhou (Section 3.3) for presence of immunity and sufficient follow-up are, not surprisingly, passed with ease. The hazard functions are clearly non-monotonic. In the case of cigarettes the hazard peaks at age 18 and tends to zero at age 26. In the case of cannabis it peaks between the ages of 17 and 21 years and tends to zero at about 32 years. In the case of hard drugs the hazard jumps sharply after age 15, peaks at age 18–20, after which it declines gradually until age 30.

The vast majority initiates cannabis after cigarettes, and initiates hard drugs after cannabis. In the next Section, we set the age of initiation for cigarettes at a theoretical value of 120 years¹⁵ for non-smokers, and we

¹⁵This age was selected as being beyond maximal life spans. It follows the Jewish tradition that there is an upper limit to longevity of 120 years, which derives from Moses' age when he died.

Table I. Hazard Rates for the Susceptible Population (h^*)

	Cigarettes	Cannabis	Hard drugs	
Immunity rate	0.3	0.86	0.92	0.98
15 years	0.083	0.029	0.015	0.065
Peak (age at peak)	0.350 (18)	0.141 (20)	0.078 (17)	0.162 (17)
20 years	0.276	0.141	0.017	0.103
25 years	0.070	0.083	0.014	0.117
30 years	0.050	0.0273	0.007	0.078

define a dummy variable, which assumes a value of unity for non-smokers and zero otherwise. In this way non-smokers, who by definition do not report an age for cigarette initiation, are not treated as missing data.¹⁶

The hazard functions in Fig. 1–3 refer to the population as a whole, i.e. both the immune and the susceptible. In Table I, we report estimates of h^* , the hazard function for the susceptible, as derived from Eq. (4). These are estimates because they are based on the proportion of the population that is estimated to be immune. This parameter is estimated in Section 5 for cannabis and Section 6 for hard drugs. For illustrative purposes the immunity rate is assumed to be 0.3 in the case of cigarettes.

Table I shows that the susceptible hazard rates are obviously much greater than their population counterparts. However, it turns out that, like their unconditional counterparts, the susceptible hazard rates are non-monotonic; the hazard initially increases and subsequently decreases. The final column in Table I shows that if the immunity rate in the case of hard drugs is high the susceptible hazard rate may be bi-modal. This reflects the flatness in the underlying hazard rate between the ages of 18 and 25 years that may be seen in Fig. 3.

5. RESULTS FOR CANNABIS

In view of the evidence in Fig. 1–3, where the survival rates bottom out at levels which clearly exceed zero, we see no point in reporting results where everybody is assumed to be susceptible. Indeed, we report below the results of statistical tests that overwhelmingly reject the standard single population model in favor of the split population model. We begin in Section 5.1 with LTSMs that are fully parametric. Thereafter in Section 5.2 we report results

¹⁶In Beenstock and Rahav (2002) we condition separately on smokers and non-smokers to identify the causal effect of cigarettes on the subsequent drug initiation. Here, however, we aggregate smokers and non-smokers because our present concern is with conditioning upon the determinants of susceptibility.

obtained using proportional hazards. Finally, in Section 5.3 we discuss the treatment of unobserved heterogeneity among the susceptible population.

5.1. Parametric LTSMs

There are four interrelated specification problems to be addressed. The first concerns the specification of the survival and hazard functions for the susceptible population. Fig. 2 for h , and Table I for h^* clearly indicate that the hazard function is non-monotonic, which apparently rules out the Weibull and exponential distributions, and suggests that the lognormal and log-logistic distributions may be suitable candidates. Since the log-logistic is more symmetrical than the lognormal, the shape of the hazard function in Fig. 2 points to the log-logistic as the main candidate.

The second specification problem concerns the distribution of immunity or susceptibility. We may either assume that immunity is normally distributed, in which case the probit model is appropriate, or we may assume that immunity has a logistical distribution, in which it should be modeled as a logit model. We have no priors about the distribution of immunity in the population.

The third specification problem concerns the choice of covariates (X) in the hazard function for the susceptible, and the fourth specification problem concerns the choice of covariates (Z) in the immunity model. The multi-dimensional nature of the specification problem means that the number of potential models is large. Space prevents us from reporting all or even many of our findings.

Not surprisingly, we found that the Weibull and exponential distributions were empirically unsuitable. We also found that for any given choice of X and Z similar results were obtained regardless of the choice of the specification of the hazard and immunity distributions, provided that the former was lognormal or log-logistic and the latter was normal or logistical. To narrow down the choice of the X and Z variables we adopted a backward step-wise procedure. In the absence of strong theoretical priors about the choice of these variables, the initial list of these variables was quite unrestricted. Indeed our only prior was that the earlier initiation of cigarette smoking was likely to induce the earlier initiation of cannabis. The number of X variables turned out to be quite limited in comparison to the number of Z variables.

In Table II, we report estimates in which immunity is assumed to be normally distributed, while the initiation density is lognormal. Experimentation using a backward step-wise procedure revealed that while immunity depends upon a wide range of covariates listed in Table II, age at initiation depended upon only two variables: age at cigarette initiation and

Table II. Parametric Mixture Model for Cannabis Immunity and Initiation

Variable	Coefficient	Standard error
<i>Hazard model:</i>		
Intercept	2.51	0.038
Non-smoker	1.03	0.042
Age at cigarette initiation	0.031	0.002
<i>Immunity model</i>		
Intercept	0.479	0.120
Asia	0.553 (1.210)	0.201
Balkan	0.688 (1.254)	0.087
Eastern Europe	0.458 (1.177)	0.093
Israel	0.229 (1.091)	0.092
Middle East	0.749 (1.273)	0.089
North Africa	0.731 (1.267)	0.085
Pub Frequency	-0.229 (0.890) ^a	0.017
Religious observance 1 + 2	0.376 (1.146)	0.059
Female	0.386 (1.150)	0.041
Israel-born	0.193 (1.076)	0.051

Log likelihood = -2,650 $\sigma = 0.243$ $N = 12,440$ Estimated using Limdep 7.0 with lognormal survival distribution and normal (probit) immunity distribution. The estimated rate of susceptibility is 0.141 with standard error of 0.00454. Goodness-of-fit: order correlation $r = 0.9941$. Risk ratio reported in parentheses.

^aRisk ratio calculated at mean of data. Reference group: foreign born irreligious men, whose fathers were born in North and South America, or W. Europe.

smoking status. A positive coefficient in the susceptibility model in fact implies lower susceptibility or greater immunity. For example, women are less susceptible and therefore more immune, while pub frequenters are more susceptible to cannabis. We suggest that women have an innate immunity since gender is fixed, whereas pub frequenters have an acquired susceptibility because pub attendance is not fixed. Table II indicates that innate immunity also depends upon origin (e.g. Israeli born are less susceptible) and father's origin (e.g. father born in Israel less susceptible). Finally, susceptibility is lower among the more religious. We cannot say whether this immunity is acquired or not because we do not know whether religiosity has been acquired or inherited. A notable absentium from the susceptibility/immunity model in Table II is education. It does not appear to be the case that susceptibility and education are related.

Whereas susceptibility depends upon a range of variables, the initiation process depends only on cigarette-related variables. People who initiate cigarette smoking later are likely to initiate cannabis later. Also non-smokers initiate cannabis later. Note that according to Table II there is no overlap at all between the Z variables in the susceptibility model and the X variables in the initiation model. The factors affecting initiation seem to be separate from those affecting susceptibility and immunity.

Table III. Sensitivity Analysis for Parametric LTSMs

	Initiation distribution	Susceptibility distribution	Log likelihood	Susceptibility rate
1. (Table II)	Normal	Lognormal	-2650	0.14061
2.	Logistic	Lognormal	-2654	0.14174
3.	Normal	Log-logistic	-2586	0.14234
4.	Logistic	Log-logistic	-2592	0.14420

The correlation coefficient between the predicted and KM orders of failure is 0.9941. This looks impressive, but because the sample size is very large the critical value for r is greater than about 0.9927.¹⁷ The estimated rate of susceptibility (reported below Table II) is about 14%, which means that if the survival function in Fig. 2 were extended, it would bottom out at about 0.86 instead of 0.91. This model does well in predicting the order of failure, but it overpredicts susceptibility.

We have estimated the model in Table III using different parametric assumptions. Results of this sensitivity analysis are reported in Table III. When susceptibility is assumed to have a logistical distribution instead of a normal distribution there is a slight deterioration in the goodness-of-fit, as measured by the log likelihood function (from -2650 to -2654), and the parameter estimates are not affected. When the initiation density is assumed to be log-logistic instead of normal, and susceptibility is assumed to be normally distributed the log likelihood function is -2586, suggesting that the log-logistic fits the data better than the lognormal. However, the parameter estimates are not affected. Indeed, a Hausman test shows that the parameter estimates are not significantly different for all the models reported in Table III. Also, the estimated rates of susceptibility are very similar across all the specifications. The main conclusion is that the results reported in Table II are robust to different parametric specifications.

5.2. Semi-Parametric Long-Term Survivor Model

There are two key differences between the parametric and semi-parametric mixture models. First, the goodness of fit of the latter is superior ($r = 0.9951$). No doubt this arises from the fact that the proportional hazards model has a more flexible survival function than its fully parametric rival. Second, the hazard model in Table IV is richer in covariates than its counterpart in Table II. The two models concur that age of cannabis

¹⁷Maller and Zhou (1996, Table C1) report percentiles of r when the rate of susceptibility is 0.2. The 20th percentile for heavily censored data is 0.9927 when $N = 1000$ and 0.9901 when $N = 900$. The 5th percentiles are 0.9872 and 0.9857.

Table IV. Semi-parametric Mixture Model for Cannabis Initiation and Immunity

Variable	Coefficient	Standard error
<i>Immunity Model</i>		
Intercept	0.1084	0.0987
Female	0.5067 (1.6598)	0.0423
Israel	0.3166 (1.3725)	0.0987
Middle East	0.6507 (1.9169)	0.0987
North Africa	0.640 (1.8965)	0.0953
Balkans	0.5789 (1.7841)	-0.0992
Asia	0.4244 (1.5287)	0.2127
Eastern Europe	0.5024 (1.6257)	0.1057
Religious Observance 1+2	0.5452 (1.7250)	0.0556
Education 4	0.3317 (1.3933)	0.0661
Education 5+6+7	0.5839 (1.7930)	0.0491
Pub frequency	-0.2052 (0.6661) ^a	0.0176
<i>Proportional hazard model</i>		
Middle East	-0.3102	0.0924
North Africa	-0.3691	0.0887
Balkans	-0.4119	0.0899
Religious observance 1+2	-0.2328	0.1042
Education 4	-0.4085	0.0971
Education 5+6+7	-0.2936	0.0701
Pub frequency	0.2019	0.0234
Non-smoker	-5.9452	0.2176
Age at cigarette initiation	-0.2040	0.0107

$N=12,440$ Log likelihood function = $-4,512$. Immunity is assumed to have a logistical distribution. Predicted rate of susceptibility = 0.117. Goodness-of-fit: order correlation $r=0.9951$. Risk ratio reported in parentheses.

^aRisk ratio calculated at mean of data. Software written with logit specification by Irena Vovk. Reference group: as in Table II in education group 1-3 and 8-9.

initiation varies directly with age of cigarette initiation. However, they disagree insofar as the hazard in Table IV depends upon a variety of demographic characteristics, whereas in Table II it does not.

According to Table IV, certain covariates affect both susceptibility and the hazard. For example, religious people not only have a significantly smaller susceptibility, but also initiate more slowly. The same applies to those with intermediate levels of education (4-7 on a 9-point scale) and to father's place of origin. Women have a lower susceptibility, but they do not initiate at a different rate than men. Note that, as in Table II, cigarette smoking is, by far, the most statistically significant variable affecting cannabis initiation, although it does not affect susceptibility/immunity.

Table III showed that the results are similar for log-logistic and log-normal densities. By contrast, a comparison between Tables II and IV shows that the results are sensitive to semi-parametric specifications of the density. While the variables that proved to be statistically significant in the two

Table V. Predicted and Actual Survival Rates for Cannabis

Age (years)	15	20	25	30	35	40	∞
Survival rate (Figure. 2)	0.995	0.941	0.917	0.908	0.905	0.902	
Predicted survival rate (Table II)	0.998	0.963	0.923	0.901	0.891	0.873	0.859
Predicted survival rate (Table IV)	0.996	0.947	0.920	0.910	0.899	0.891	0.883

tables are similar, their roles in explaining immunity and susceptibility are different. However, both models reach the same conclusion regarding the relationship between cigarettes and immunity. Hence this result is robust.

A further difference between Table II and Table IV is that in the latter the estimated rate of susceptibility is 11.7%, which is smaller than the 14% obtained in Tables II and III. Presumably, the more flexible specification of the hazard function in Table IV is responsible for this result.¹⁸ Indeed, the model in Table IV tracks the survival curve in Fig. 3 much more faithfully than the model in Table II. A comparison of the two models is provided in Table V. Both models initially over-predict survival and subsequently underpredict it. However, the errors of the model in Table IV in particular are quite small and acceptable, as reflected in the superior value of $r = 0.9951$. This suggests that overprediction by the model in Table IV of some 2½ percentage points in the rate of susceptibility largely results from extrapolation beyond the support of the data.

5.3. Frailty and Heterogeneity

In Section 3.5, we raised the possibility that the susceptible population may not be homogeneous, and that the degree of susceptibility may vary. We wish to emphasize that the issue of heterogeneity among the susceptible is of secondary or even minor importance when compared with the major issue of immunity and susceptibility. Because of software constraints we could only investigate the matter for the Weibull case with gamma heterogeneity. The specific question that we address here is whether the \cap -shaped hazard function observed in Fig. 2 is in fact an optical illusion resulting from heterogeneity, when the true distribution is Weibull.

It should be stressed that these models are difficult to estimate because they mix three distributions instead of two as in LTSM. These are Weibull and gamma distributions for the susceptible, and normal or logistic distributions for the immune. Convergence and matrix singularity proved to be problematic in many cases. However, these problems were mitigated when

¹⁸Kuk and Chen (1992) also report that proportional hazards better predict the rate of susceptibility.

we restricted the number of Z variables to four (father born in Israel, father born in N. Africa, religiosity and frequency of pub visits) instead of 10 as in Tables II and IV. The value of the log likelihood function obtained was -3588 , which is greatly inferior to that obtained in Table II, and the estimate of φ was not statistically significant. This means that we reject the notion that the distribution for the susceptible is Weibull with gamma heterogeneity.

6. RESULTS FOR HARD DRUGS

In this Section we report multivariate failure time models for the initiation of hard drugs, defined as the earliest age at which one or more of hard drugs (cocaine, heroin, LSD, and amphetamines) were first used. In estimating these models we followed the specification procedures described in Section 5.1. Problems of convergence and singularity proved to be more common. We confine ourselves to the parametric mixture model reported in Table VI, which mixes a log-logistical density for initiation and a normal density for immunity.¹⁹

The mixture model shows that the probability of being immune is qualitatively similar to that for cannabis. For example, the highly religious have a greater probability of immunity to hard drugs, immunity varies inversely with pub visits, and immunity is greater among the higher educated. People of Eastern European backgrounds are less likely to be immune (although not significantly so) as are people with Balkan and Middle Eastern origins. Asian and North African origin, which increased susceptibility to cannabis (Tables II and IV), do not apparently increase susceptibility to hard drugs. Finally, people surveyed in 1989 are more susceptible to hard drugs than those surveyed in 1992 and 1995. This suggests a fall in susceptibility to hard drugs during the first half of the 1990s.

Only two variables turned out to be statistically significant in the hazard model for hard drugs. Noticeably absent is age at cannabis initiation. The hazard model implies that age at hard drug initiation varies directly with age at cigarette initiation, and that non-smokers initiate later.

The order correlation coefficient for goodness-of-fit ($r = 0.981$) is lower in the case of hard drugs than it is in the case of cannabis. This is to be expected because Fig. 3 indicates that the rate of susceptibility is very low for hard drugs.²⁰ The estimated rate of susceptibility is 8.1%, which means that the survival curve in Fig. 3 would eventually bottom out at about 0.92. This is clearly an overestimate. Table VII compares

¹⁹We were unable to apply the proportional hazards methodology to the data for hard drugs.

²⁰Maller and Zhou show that when $N = 1000$ the critical value for r at the 10th percentile falls from 0.9938 to 0.9901 when the rate of susceptibility falls from 0.3. to 0.2.

Table VI. Parametric Mixture Model for Hard Drug Initiation

Variable	Coefficient	Standard error
<i>Hazard model</i>		
Intercept	1.6426	0.2250
Age at cigarette initiation	0.1345	0.0185
Non-smoker	3.2388	0.4742
Eastern Europe	-0.0310	0.6232
<i>Immunity model</i>		
Intercept	1.0585	0.1880
Israel	0.1856 (1.0736)	0.1569
Middle East	0.3468 (1.1356)	0.1508
Balkan	0.4228 (1.1638)	0.1634
Eastern Europe	0.7402 (1.2705)	0.6489
Education 4	0.3443 (1.1347)	0.1415
Education 5-7	0.4103 (1.1592)	0.1201
Religious 1-2	0.2866 (1.1128)	0.1800
Religious 3	0.2052 (1.0813)	0.1200
Pub frequency	-0.2013 (0.8480) ^a	0.0428
Survey 1989	0.2514 (1.1001)	0.1275

$\log L = -643$ $N = 12,504$ $\sigma = 0.3332$ Estimated with LIMDEP 7.0 using a log-logistic initiation distribution and normal immunity distribution. Estimated susceptibility rate = 0.081 with standard error = 0.016. Goodness-of-fit: order correlation $r = 0.9803$.

^aRisk ratio calculated at mean of data. Reference group as in Table IV. Risk ratio reported in parentheses.

the actual rate of survival with the predicted rate from the model in Table VI. It shows that the overprediction of the rate of susceptibility largely results from extrapolating beyond the support of the data. However, this problem begins to emerge beyond the age of 30, when the empirical survival curve in Fig. 3 bottoms out. An error of 0.01 at aged 40 may be small in absolute terms but is large in relative terms. More generally, cTable VI fits the hard drugs data less well than do Table IV and even Table III fit the cannabis data.

7. CONCLUSIONS

The statistical literature on drug initiation has implicitly assumed that given enough time everyone will start using drugs. By contrast, the data suggest that drugs are a minority activity; not everyone gets into

Table VII. Predicted and Actual Survival Rates for Hard Drugs

Age (years)	15	20	25	30	35	40	∞
Survival rate (Fig. 3)	0.9983	0.9917	0.9882	0.9866	0.9864	0.9862	
Predicted survival rate (Table VI)	0.9984	0.9921	0.9891	0.9851	0.9801	0.9767	0.9190

drugs. We suggest that it is more appropriate to divide the population into two groups, a sub-population who is immune to drugs and will never use them no matter how long they live, and a susceptible sub-population for whom it is a matter of time before they initiate drugs. We have used alternative models to estimate these two sub-populations. While it is clearly wrong to assume zero immunity, it is difficult to identify the two populations in the data. This seems to be particularly the case when the data are heavily censored, as they are in our case.

Long-term survivor analysis assumes that the hazard and susceptibility are stochastically independent. If people who are more susceptible to drugs also tend to initiate drugs when they are younger, this assumption will be wrong. In this case the residual error in the model for susceptibility will be positively correlated with the residual error in the hazard model for the susceptible. Ignoring this co-dependence could bias the estimates of the model. Indeed, we suspect that this lies behind our tendency to overpredict the rate of susceptibility, especially in the case of hard drugs. However, the statistical methodology for taking account of this co-dependence has not yet been developed. We also suspect that when the susceptibility rate is very low and approaches zero, as is the case with drug use data in general and hard drugs in particular, it becomes very difficult to predict the rate of susceptibility. A prediction error of say 5 percentage points looks more glaring when the rate of susceptibility is 2% than when it is 50%.

We have tried to distinguish between immunity and susceptibility in the drug initiation process. This distinction is important because certain interventions may not affect drug use although they may delay initiation, while others may affect drug use without delaying initiation. The dichotomy between immunity and susceptibility is an important conceptual breakthrough in the study of illicit drug use. We have suggested alternative statistical methodologies for identifying the determinants of immunity, and for explaining the drug initiation hazard for the susceptible. Since this is a new issue on the research agenda of illicit drug use, we wish to summarize our findings with more than the usual modicum of modesty.

Cigarettes affect the timing of drug use initiation, both for cannabis and hard drugs, but they do not seem to have any relevance for immunity. This result is robust and is not sensitive to the method of estimation, especially in the case of cannabis. It implies that anti-smoking campaigns may merely delay drug initiation; they do not affect immunity. While the estimated relationship between cigarettes and cannabis is consistent with Gateway Theory, the estimated relationship between cannabis and hard drugs is not. In the present sample, cannabis use is not a good predictor of hard drug

initiation. While this may be partly due to the small number of hard drug users in the sample, this finding highlights the need for caution in using Gateway Theory.²¹ On the other hand, cigarettes are a gateway to hard drugs in Israel. Since most heroin users in Israel consume it by smoking, it is quite possible that smoking provides heroin users with appropriate background experience.

The crime–delinquency model of drug use initiation is not, on the whole, supported by our findings, especially for cannabis. People of Middle Eastern and lower socio-economic backgrounds, which together provide good predictors of delinquency in Israel, are more immune to cannabis and hard drug initiation. In the case of cannabis higher educated Ashkenazi Jews are more susceptible. These results do not concur with models of deviant behavior (Dohrenwend *et al.* 1992, and Srole *et al.* 1962), and suggest that the deviant behavior model is not necessarily suitable to drug initiation in Israel. Instead they support the hypothesis that this type of deviance is least prevalent among the intermediate strata (Reckless, 1967); both in the case of cannabis and hard drugs those with tertiary and primary levels of education were most susceptible.

Bahr *et al.* (1998) have drawn attention to the importance of religious devotion in the determination of drug use. Theirs was a Christian context, ours is a Jewish context. Although Judaism and Christianity differ in dogma and the relative weight given to belief and ritual, the results for drug use seem to be similar. However, the present study allows us to make a finer distinction. We suggest that religious devotion increases immunity to drugs rather delay initiation. Moreover, because the sample consists entirely of Jews, we suggest that it is religious devotion *per se* which is associated with immunity from drugs rather than community belonging.

Our results show that immunity is not randomly distributed across the population. Some groups have a higher propensity to be immune than others. For example, Table II shows that women tend to have a higher rate of immunity to cannabis than men. An interesting question is whether immunity is determined at birth and is natural, or whether it can be acquired. Insofar as the variables in the immunity model are fixed at birth (such as sex, parental origin) immunity cannot be acquired. Insofar as the variables in the immunity model are not fixed at birth (such as education and religious practice) then immunity may have an acquired component. Future work may usefully distinguish between acquired and natural immunity.

²¹See also Beenstock and Rahav (2002) who distinguish between the gateway effect as a causal phenomenon and as an incidental phenomenon.

ACKNOWLEDGMENTS

This study was partly funded by the Israel Science Foundation. We wish to thank the Israel Anti-Drugs Authority for providing the data, Irena Vovk and Michael Hvoshnyansky for their excellent research assistance, and four referees for their helpful remarks.

REFERENCES

- Bahr, S. J., Maugham, S. L., Marcos, A. C., and Li, B. (1998). Family, religion, and the risk of adolescent drug use. *J. Marriage Fam.* 60: 979–992.
- Beenstock, M., and Rahav, G. (2002). Testing Gateway Theory: The effect of cigarette prices on the consumption of cigarettes and illicit drugs in Israel. *J. Health Econ.* 21: 679–698.
- Chilcoat, H. D., and Anthony, J. C. (1996). Impact of parent monitoring on initiation of drug use through late childhood. *J. Amn. Acad. Child Psych.* 35: 91–100.
- Chilcoat, H. D., and Schutz, C. G. (1996). Age-specific patterns of hallucinogen use in the US population: An analysis using generalized additive models. *Drug Alcohol Depen.* 43: 143–153.
- De Wit, M. L., Embree, B. G., and De Wit, D. (1999). Determinants of the risk and timing of alcohol and illicit drug use onset among natives and non-natives: Similarities and differences in family attachment. *Social Biol.* 46: 100–121.
- Dohrenwend, B. P., Levav, I., Shrout, P. E., Schwartz, S., Naveh, G., Link, B. G., Skodol, A. E., and Stueve, A. (1992). Socioeconomic status and psychiatric disorders: The causation-selection issue. *Science* 255: 946–952.
- Douglas, S., and Hariharan, G. (1994). The hazard of starting smoking: Estimates from a split population duration model. *J. Health Econ.* 13: 213–230.
- Escarela, G., Francis, B., and Soothill, K. (2000). Competing risks, persistence, and desistance in analyzing recidivism. *J. Quant. Criminol.* 16: 385–414.
- Fergusson, D. M., and Horwood, L.J., (1997). Early onset of cannabis use and psychological adjustment in young adults. *Addiction* 92(3): 279–296.
- Greene, W. A. (2000). *Econometric Analysis*. 4th edn. Macmillan, New York.
- Heckman, J., and Singer, B. (1984). Econometric duration analysis. *J. Econom.* 24: 63–132.
- Kandel, D. B., and Logan, J. A. (1984). Patterns of drug use from adolescence to young adulthood: I. periods of risk for initiation, continued use, and discontinuation. *Am. J. Pub. Health* 74: 660–666.
- Knupfer, G., and Room, R. (1967). Drinking patterns and attitudes of Irish, Jewish, and white Protestant American men. *Quart. J. Alcohol Stud.* 28: 676–699.
- Kuk, A. Y. C., and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 79: 531–541.
- Larson, M. G., and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Appl. Stat* 34: 201–211.
- Luthar, S. S., Anton, S. F., Merikangas, K. R., and Rounsaville, B. (1992). Vulnerability to substance abuse and psychopathology among siblings of opioid users. *J. Nerv. Ment. Dis.* 180: 153–161.
- Maller, R., and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*, John Wiley.
- Rahav, G. (1998). Juvenile delinquency. In Friedmann, R. R. (ed), *Crime and Criminal Justice in Israel*, University of New York Press, New York.

- Rahav, G., Teichman, M. Gil, R., and Rosenbloom, I. (1996). *The Use of Psychoactive Substances among Residents of Israel 1995, Epidemiological Study III*. Israel Anti-Drug Authority, Jerusalem (Hebrew).
- Reckless, W. C. (1967). *The Crime Problem*. Appleton – Century – Crofts, New York.
- Schmidt, P., and Witte, A. (1989). Predicting criminal recidivism using split sample population survival time models. *J. Econom.* 40: 141–159.
- Segal, B. (1991). Adolescent initiation into drug-taking behavior: Comparisons over a 5-year interval. *Int. J. Addict.* 26: 267–279.
- Srole, L., Langner, T. S., Michael, S. T., Opler, M. K., and Renie, T. A. C. (1962). *Mental Health in the Metropolis: The Midtown Study*. McGraw Hill, New York.
- Tsibel, N. (2000). *Event History Analysis Accounting for Measurement Errors and Missing Data*, Phd dissertation, Hebrew University of Jerusalem, Israel.
- Turnbull, J. E., George, L. K., Landerman, R. Swatz, M. S. *et al.* (1990). Social outcomes related to age of onset of psychiatric disorders. *J. Consult. Clin. Psychiat.* 58: 832–839.
- Willens, T. E., Biederman, J., Mick, E., Pharone, S. V., and Spencer, T. (1997). Attention deficit hyperactivity disorder (ADHD) is associated with early onset of substance abuse. *J. Nerv. Ment. Dis.* 185: 475–482.
- Yamaguchi, K., and Kandel, D. B. (1984). Patterns of drug use from adolescence to young adulthood: III predictors of progression. *Am. Public Health* 74: 673–82.
- Yu, J., and Williford, W. R. (1992). The age of alcohol onset and alcohol, cigarette, and marijuana use patterns: An analysis of drug use progression of young adults in New York state. *Int. J. Addict.* 27: 1313–23.