UNITED NATIONS INTERNATIONAL DRUG CONTROL PROGRAMME

BULLETIN ON NARCOTICS

Volume LIII, Nos. 1 and 2, 2001

Dynamic drug policy: Understanding and controlling drug epidemics

UNITED NATIONS

UNITED NATIONS INTERNATIONAL DRUG CONTROL PROGRAMME
Vienna

BULLETIN ON NARCOTICS

Volume LIII, Nos. 1 and 2, 2001

Dynamic drug policy: Understanding and controlling drug epidemics

UNITED NATIONS
New York, 2002

UNITED NATIONS PUBLICATION
Sales No. E.02.XI.6
ISBN 92-1-148147-3
ISSN 0007-523X

PREFACE

The *Bulletin on Narcotics* is designed to provide information on developments in drug control at the local, national, regional and international levels that would benefit the international community. It is a United Nations publication that is available in Arabic, Chinese, English, French, Russian and Spanish.

Individuals and organizations are invited by the Editor to contribute articles to the *Bulletin* dealing with policies, approaches, measures and developments (theoretical and/or practical) relating to various aspects of the drug control effort. Of particular interest are the results of research, studies and practical experience that would provide useful information for policy makers, practitioners and experts, as well as the public at large.

The present issue of the *Bulletin* contains papers that were originally prepared for a symposium on the theme "Dynamic drug policy: understanding and controlling drug epidemics" that was organized by the United Nations International Drug Control Programme and the Technical University of Vienna from 22 to 24 May 2000.

Editorial policy and guidelines for publication

All manuscripts submitted for publication in the *Bulletin* should constitute original and scholarly work that has not been published elsewhere or is not being submitted simultaneously for publication elsewhere. The work should be of relatively high professional calibre in order to meet the requirements of a United Nations technical publication. Contributors are kindly asked to exercise discretion in the content of manuscripts so as to exclude any critical judgement of a particular national or regional situation.

The preferred mode of transmission of manuscripts is in Word format. Each submitted manuscript should consist of an original hard copy and a 3.5" diskette, in Word for the text and Excel or Lotus for charts and tables, in any of the six official languages of the United Nations. The manuscript should be accompanied by an abstract of approximately 200 words and by a complete set of references numbered in the order of their appearance in the text. The manuscript should be between 10 and 20 double-spaced typewritten pages, including tables, figures and references. Tables should be self-explanatory and should supplement, not duplicate, information provided in the text.

Manuscripts, together with brief curricula vitae of their authors, should be addressed to the Editor, *Bulletin on Narcotics,* United Nations International Drug Control Programme, P.O. Box 500, A-1400 Vienna, Austria. A transmittal letter should designate one author as correspondent and include his or her complete address, telephone number, facsimile number and, if available, e-mail address. Unpublished manuscripts will be returned to the authors; however, the United Nations cannot be held responsible for loss.

Views expressed in signed articles published in the *Bulletin* are those of the authors and do not necessarily reflect those of the United Nations Secretariat. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat concerning the legal status of any country, territory, city or area, or its authorities, or concerning the delimitation of any frontiers or boundaries.

Reprints, purchases and subscriptions

All issues of the Bulletin (from vol. I, No. 1 (1949), to the present issue) are available on the home page of the United Nations International Drug Control Programme, at http://www.undcp.org.

The following special issues of the Bulletin are also available as United Nations publications:

1990

Emerging directions and trends in drug abuse control (vol. XLII, No. 1)

1991

Involvement of intergovernmental and non-governmental organizations in matters of drug abuse control (vol. XLIII, No. 1)

1992

The role of law enforcement agencies in drug abuse control (vol. XLIV, No. 1)
The environmental impact of drug abuse (vol. XLIV, No. 2)

1993

Policy issues relating to drug abuse and the human immunodeficiency virus (HIV) (vol. XLV, No. 1)
Drug testing in the workplace (vol. XLV, No. 2)

1994

The family and drug abuse (vol. XLVI, No. 1)
General issue on drug abuse (vol. XLVI, No. 2)

1995

Special issue on gender and drug abuse (vol. XLVII, Nos. 1 and 2)

1996

Special issue on rapid assessment of drug abuse (vol. XLVIII, Nos. 1 and 2)

1997 and 1998

Double issue on cannabis: recent developments (vol. XLIX, Nos. 1 and 2 and vol. L, Nos. 1 and 2)

1999

Occasional papers (vol. LI, Nos. 1 and 2)

2000

Economic and social costs of substance abuse vol. LII, Nos. 1 and 2)

Requests for permission to reprint signed material should be addressed to the Secretary of the Publications Board, United Nations, New York, New York 10017.

Correspondence regarding the purchase of copies of and subscriptions to the *Bulletin* should be addressed as follows:

For Asia, North America, Oceania and South America:

The Chief
Sales and Marketing Office in New York
United Nations Publications
United Nations Headquarters
New York, New York 10017
United States of America

For Africa, Europe and the Middle East

The Chief
Sales and Marketing Office in Geneva
United Nations Publications
United Nations Office at Geneva
Palais des Nations
CH-1211 Geneva 10
Switzerland

CONTENTS

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

The need for dynamic models of drug markets*

P. REUTER

Professor, School of Public Affairs, University of Maryland, College Park, Maryland, and Senior Fellow, RAND Drug Policy Research Center, Santa Monica, California, United States of America

ABSTRACT

Markets for cocaine and heroin are created out of epidemics and are best analysed in terms of dynamic models rather than comparative statics. The author presents the following three problems requiring dynamic models and sketches a potential approach to each one: (a) *why prices have declined so persistently in the face of intensified enforcement;* (b) *informing the choice between buyer and seller targeting for enforcement; and* (c) *explaining the extremely high variability of retail prices in such markets. For the first two problems, critical issues are the modelling of the path of sources of earnings (such as other crime or legitimate activities) and the drug-selling labour supply as consumption becomes dominated by drug users with increasingly long criminal histories and fewer legitimate employment opportunities. For the third problem, models need to take into account both the difficulty of ascertaining quality at the time of retail purchase and the instability of use and sale opportunities within those markets.*

Introduction

Illicit cocaine and heroin markets in the United States of America** appear to be different from legal markets in a number of ways: high levels of violence; rapid turnover of participants; the association, at the individual level, of frequent use and selling; and the large variation of prices and quality in narrowly defined geographic markets at a given point in time. These are not universal characteristics of other illegal markets in the United States. For example, the market for illegal gambling services in the 1970s was characterized by stable participation and uniform prices; bookmakers tended to be heavy bettors themselves but that was not true of

*The research reported in the present article was supported by the National Institute of Justice of the United States Department of Justice. The author is grateful to Jonathan Caulkins for his helpful comments.

**The cocaine and heroin markets are the most important illicit drug markets in the United States, in terms of both the income generated and the consequent social costs.

numbers operators. Furthermore, the gambling markets were subject to light enforcement [1].

Moreover, violence and instability, for example, are not fixed characteristics of the market for cocaine or heroin but may be sensitive to characteristics of users and sellers (such as age and relative number of light and heavy users) that change in a systematic fashion, as well as to the intensity and form of law enforcement. That suggests the need for dynamic models, which take into account the fact that those markets evolve in the course of an epidemic and the multidimensional responses of those markets to policy interventions. Yet the model of risks and prices [2] most used in empirical policy models [3] is a strictly comparative static model, aiming only to describe the long-run adjustment of a market whose principal distinguishing characteristics are the centrality of enforcement risk and violence by other participants in the determination of prices.

The present article outlines three topics that go beyond the framework of risks and prices and seem to lend themselves to dynamic models. They each represent a class of problems that are explanatory, policy-aiding and conceptual. The first topic focuses on an explanation for the persistent long-term decline in prices as the illicit cocaine and heroin markets develop. The second topic addresses whether enforcement resources should be allocated over the epidemic between customer sanctions and seller sanctions. The focus of the third topic is the determinants of variations in price and purity. Each topic requires its own modelling approach. The discussion below begins with a synopsis of the evolution of the cocaine market in the United States.

The evolution of drug markets

Drug epidemics, at least for expensive* and dependency-creating drugs,** are characterized by sharp peaks in population incidence rates followed, with a lag, by a plateauing at a new high in the number of dependent users. The pattern reflects the fact that a portion of new users become dependent within a few years, that incidence is partly driven by the extent of perceived problematic use and that exit from dependence is slow. Everingham and Rydell [5] offer a now classic representation of that phenomenon, while Behrens and others [6, 7] have explored the dynamics in more detail. Drug markets vary over the course of a drug epidemic in the ratio of heavy users to light users, the mean age of users and, in a predictable

*Price is usually measured per unit weight. Thus, heroin, selling at 500-1,000 United States dollars per pure gram, is more expensive than cocaine, selling at approximately US$ 100 per gram. However, in terms of dose (that is, the amount taken in a typical session in which a drug is used), there may be little difference. Yet another possible metric, consistent with the rational addiction model of Becker and Murphy [4], is per annum expenditure or expected lifetime expenditures: cocaine may be more expensive in the first of those and substantially cheaper in the latter.

**Marijuana is both dependency-creating (with about 10 per cent of users self-reporting dependence at some time in their lives) and expensive (US$ 5-10 per gram); however, the dependency appears to be short-lived, that is, it lasts rarely more than a few years and the annual expenditures are much smaller for marijuana than for cocaine or heroin.

fashion, the opportunity cost of sellers' time. Those in turn affect the level of property crime generated by drug use.*

In the early stages of the cocaine epidemic in the United States, drug users were not predominantly poor. The image of the drug was relatively benign, its dangers were little known and its attractions were great. Most users were inexperienced and did not at that time consume large quantities or suffer significant problems. Low-income users could earn substantial incomes selling to users who are not poor.** Such conditions are likely to be common in the early stages of drug epidemics in which the drugs are not well known to the population.

In the late 1980s, frequent users made up a much larger fraction of all cocaine users and accounted for a larger fraction of total cocaine consumption. Cocaine users were then poorer and had acquired both a criminal history and a record of treatment. More educated cocaine users were likely to have responded to messages about the adverse consequences of the drug and to experience better outcomes in treatment. Evidence from the National Household Survey on Drug Abuse shows that the negative correlation of current cocaine use (that is, use in the previous month) and education increased substantially after 1985 [9].

The continued decline in prices

Law enforcement responded to increased cocaine use with a lag, as would be expected with any problem that emerges rapidly. Reuter [10] suggests that the stringency of law enforcement by various metrics (arrests or incarcerations as the numerator; number of users, transactions or sellers as the denominator) may have fallen during the first half of the 1980s as the market expanded. The stringency of law enforcement then increased after 1985 as the market stabilized, at least in terms of the quantity consumed. The early decline in cocaine prices is consistent with enforcement swamping [11],*** as well as with the framework of risks and prices, with "learning by doing" reinforcing the effect of lower pressure from law enforcement [12]. In both analytic frames, it is difficult to account for the continued sharp decline until 1989 and the more modest declines of the 1990s, since the stringency of law enforcement rose sharply over an extended period.

One possible explanation for the time pattern of prices focuses on the changing income of cocaine users as well as on risks. Occasional and affluent users may have a low price elasticity, since cocaine expenditures account for a small share of their total expenditures. On the other hand, they may be very sensitive to any

*The issue of violence warrants separate treatment. Discussions of violence in drug markets focus on the incentives of sellers as independent agents, either resulting from competition (seeking territorial control) or transactional uncertainty (disagreement about the appropriate quantity of drugs or money). Intra-organizational violence, however, may also have important consequences for the markets. There are two forms of violence generated within organizations: *(a)* disciplinary violence directed by managers against agents who can either defect with goods or serve as informants against more senior figures; and *(b)* successional violence, the means by which a junior member of the organization may attain leadership.

**Retailers were reported to be earning US$ 30 per hour in 1988 in the Washington, D.C., market [8].

***Enforcement swamping grows out of the framework of risks and prices. If enforcement risk is the dominant source of costs for drug dealers, then increased volume, other things being equal, lowers cost; that generates the perverse phenomenon of a downward sloping supply curve.

increase in risks associated with purchase, such as use of "sell and bust" operations,* or simply higher arrest risks. The opposite patterns may be found with poor dependent users: high price elasticity** (reflecting the dominance of drugs in their consumption baskets)*** and little sensitivity to changes in arrest and other purchase risks.

Everingham and Rydell [5] developed a model with light and heavy cocaine users who differed only in their intensity of use. The outline above suggests a model in which users are classified not simply as light and heavy users in terms of consumption per unit time, but also by income or education. Each group has a potentially different elasticity of demand with respect to price and to other kinds of risks. The simplest useful version of the model has three demand components: the light user group is divided into poor and non-poor, while those in the heavy user group are all poor. The segments are linked in that it is the non-poor buyers whose purchases finance consumption by low-income users. The decline in the demand curve of the non-poor then generates a similar downward shift in the demand by poor heavy users, since their income falls.

With a fixed supply curve, reduced consumption and lower prices are generated; however, what is observed is relatively flat consumption (quantity) and declining prices [15]. The explanation for this may lie in the dynamics of the supply curve. Poor frequent users have three sources of income: legitimate work, non-drug crime and selling to non-poor users.**** The predicted declines in their income from legitimate work may be sufficient to explain the downward shift in the supply curve, given that the marginal return to property crime is presumably declining for any individual.

Turning such verbal conjecture into a formal model that takes account of the changing population of users is a major undertaking. It will require not only many of the heuristics underlying the models developed at RAND to deal with links between demand and supply [13], but also a more complex dynamic structure for those links. For example, both the probability of moving from light use to abstinence and the elasticity of demand may need to be modelled as a function of user income. However such a model may help to generate a parsimonious account of one of the most puzzling aspects of drug markets in the last decade.

Choosing between buyers and retailers as enforcement targets

The outline of market dynamics may also help in developing a model for guiding decisions about allocations between buyer and seller sanctions. It is widely

*"Sell and bust" reverse the usual "buy and bust" tactics. Instead of using undercover agents or informants to apprehend sellers through controlled buys, the police pose as sellers and catch intended buyers.

**Caulkins and Reuter [13] report a number of studies with price elasticities of one or more in absolute value. The highest values were obtained with a data system associated with heavy users (entitled "Drug Use Forecasting"). Lower values were obtained with data systems associated with broader populations of users (the National Household Survey on Drug Abuse and the data system entitled "Monitoring the Future", sponsored by the United States National Institute on Drug Abuse).

***Needle [14] reports on the high share of income going to cocaine purchases.

****Clearly, the poor sell to each other as well; however, for the group as a whole, that is merely redistributive.

assumed that sellers are more culpable, hence the heavier sanctions for those convicted of selling rather than buying or using and the greater intensity of law enforcement aimed at sellers. However, it is easy to argue for a reverse hierarchy of culpability when the seller is an impoverished dependent user and the buyer is non-dependent and non-poor.* Furthermore, sellers can be replaced; buyers cannot. There is regular reporting, for example, of "sell and bust" operations, aimed at closing down specific geographic markets.

The change in the population of frequent users may have increased the share of total expenditures derived from crime other than drug selling through three mechanisms: *(a)* declines in the fraction of users who are non-poor light users; *(b)* declining employment opportunities for frequent users with low levels of education, as their addiction and criminal histories made them less attractive to employers; and *(c)* decreased willingness of families and friends to provide monetary or any other kind of support. Since the market had contracted in terms of the number of users, who were on average poorer than they had been in the past, potential drug market earnings of the growing pool of dependent users declined.

It is assumed that the goal of drug law enforcement is to minimize the total harm caused by drug use and drug control. A first approximation can be represented as a function of the number of light users, the number of heavy users and drug-related crime other than drug selling or drug use: crime is an increasing function of the number of heavy users and the ratio of the number of heavy users to the number of light users.

Each light user is at risk of becoming a heavy user. Again, the assumption is that light users have low price elasticity but are highly sensitive to the risk of arrest. A shift away from seller enforcement, assumed to raise prices, towards buyer enforcement will reduce the number of light users in the next period; the notional budget here is a fixed total number of arrests. The change in ratio of the number of heavy users to the number of light users in that next period will be a function of two parameters: *(a)* the probability that a light user will become a heavy user and *(b)* the increased shift towards abstinence resulting from the shift in arrest activity. Crime will be affected by the same two parameters.

Those parameters are not fixed: they change in systematic ways in the course of an epidemic. For example, the shift away from light drug use towards abstinence may be low in the earliest part of the epidemic, increase rapidly and then decline, since later recruits have more knowledge of the drug than their predecessors. The optimal allocation of effort may also vary, with a larger share of the budget going to "sell and bust" operations in the early stages of the epidemic when there are many non-criminal light users at risk of becoming criminal heavy users. Later in the epidemic, the primary effect on crime may be changes in the ratio of the number of light users to the number of heavy users, leading to a shift towards conventional seller enforcement.

*It is interesting to note that for another illegal market, prostitution, Sweden has recently instituted legal reforms that shift all criminal penalties from the prostitute to the customer.

This is a highly simplified model. The objective function for enforcement decisions is broader than crime and prevalence; for example, buyer-oriented enforcement may generate markets with lower disorder. Similarly, the budget representation as a fixed number of arrests is a convenient simplification; budgets are more likely to be financial and the costs of the two types of arrests will differ. However, decisions about enforcement strategies, if they are to be made on more than an impressionistic basis, require that kind of dynamic modelling.

Variation of prices and purity

A prominent feature of drug markets is the extraordinary variation in price and purity. Weatherburn and Lind [16] report the most detailed data on prices from a single market for heroin (in a suburban area of Sydney, Australia): about 300 observations over a two-year period. The price per gram in a two-year period ranged from 118 to 11,667 Australian dollars. Even the average price per pure gram for a "fortnight" (a two-week period) showed dramatic changes, for example, collapsing from about $A 6,000 (in "fortnight 7") to $A 2,000 (in "fortnight 11"). Similar variation appears to characterize prices in the United States, though no similarly detailed analysis of a specific local market has been published.* Models of user and seller behaviour should accommodate the enormous uncertainty in the cost and purity of transactions, but the paradigm of risks and prices focuses on expected value calculations.

This price variation is made possible by an odd feature of the market: fixed prices but variable quantity that cannot be readily assessed by the buyer at the time of purchase. A dime bag of heroin always costs US$10; it may always weigh 0.1 gram, but whether it contains 50 milligrams or 2 milligrams of pure heroin cannot be ascertained at the time of purchase. It may not be ascertainable even after consumption. The diluents may mimic the drug's effects and the user may have only a general notion of how much of the drug was actually consumed. The user will make an assessment of the quality of the experience, but it will have an uncertain relationship to the actual quantity consumed.

If variation is anticipated, users can adapt. One mode of adaptation is to identify sellers who provide predictable quantities of the drug. Such sellers may charge higher average prices per pure milligram of heroin as a consequence. Prices may still seem confusing then because two classes of sellers emerge: one selling at a high mean price with low variance and the other selling at a low mean price with a high variance. The basis for segregation may include the time of day of the transaction, the circumstances and/or the place. For example, late-night purchases may be lower-priced as dealers seek to unload stocks because the market is thinning out and their risks per unit time are increasing. Purchases in which buyer

*The data from the System to Retrieve Information from Drug Evidence (STRIDE) allow only analysis at the city level, which is much broader. The Domestic Monitor Program of the Drug Enforcement Administration of the United States Department of Justice, providing data only for heroin, contains the geographic identifiers that would allow analysis of more localized markets, but rarely provides more than about 10 observations per quarter.

and seller are known to each other may have more predictable prices, in part to avoid retaliation by disappointed buyer against seller.*

Buyers' attitudes towards risk and buyer experience would affect the distribution between the two. Ethnographic reports of continual information exchange among users about the quality of different dealers' drugs suggest the existence of such heterogeneity. For example, Simon and Burns [17] describe how some heroin dealers from Baltimore distribute samples early in the day to a few experienced users in order to increase demand for that day's product.

There are more complications. Buyers and sellers frequently leave the market unexpectedly, as the result of arrest, incarcerations, injury or death. In such circumstances, it is not clear whether the optimal strategy is to develop a strong reputation for reliable selling. In strategic games of repeated interaction, whether it is optimal to cooperate (that is, sell high-quality drugs) or defect (that is, sell low-quality drugs) depends on the probability that the game terminates after any given move [18]. A seller does not want to stop selling with any reputational asset remaining: reputations may not be well disseminated because of "churning" (movement or change) among buyers. This does not imply that the optimal strategy is to defraud every customer all the time. If a dealer knows with certainty that he will leave the drug market immediately after the next sale, it would be to his advantage to cheat the next customer. A strategic investment decision needs to be made about how much reputation to maintain.

Varying attitudes towards risk, as well as variations in assessing those risks, will lead to price dispersion. That may drive the market to an equilibrium in terms of low price and high variance, with no sellers choosing to provide quantity that is predictable. Tougher enforcement may shift that equilibrium to still lower price and higher variability, since the dealer, with a lower probability of being able to reap the returns on his reputation for, say, the next three months, will choose to take advantage of more opportunities to defraud buyers.**

Although limited, the empirical literature shows that buyers of cocaine and heroin usually have substantially more than one supplier [19, 20]. That is likely to be an optimal strategy, given the turnover of suppliers and the fact that so many of them work part-time. For example, perhaps as many as half of all heroin retailers are arrested in the course of a year; injury and their own drug habit may make them unavailable at other times. That diversification of sources may limit the feasibility of establishing local monopolies, since customers will always be seeking diversification of sources to ensure reliable supplies.

The latter point requires elaboration. The customer may seek to avoid dependence on a single seller because sellers, not organizations, are incarcerated or injured. The customer's additional sellers may work for the same organization in the neighbourhood, so having multiple sources is not incompatible with buying from a single retail organization. However, the retail agents themselves also seek

*In recent reports of price and purity, it was noted that the heroin market was characterized by a bimodal distribution of pure gram prices [15].

**The discount rate of a given user or seller may also vary over time. When the need for drugs is urgent, discounted prices for drugs may be hyperbolic, with the buyer or seller sacrificing all future gains in order to obtain cocaine or heroin immediately. That would also generate price and purity dispersion.

diversification of sources. Incarceration of their principal forces them to seek supplies elsewhere, mitigating against broad retail monopolies.

Prices (per pure gram) are determined competitively. If buyers and sellers were anonymous and the buyer could not ascertain the content of purchases at the time of the transaction, then the profit-maximizing strategy would be to defraud every purchaser by providing zero-content bundles. The market would collapse, since buyers would seek alternative sources of intoxication.

The market exists; thus, the model is too simple. Even those who buy in truly anonymous transactions and who obviously pose no threat to the seller (for example, the suburban user purchasing nervously in a drive-by inner-city market) receive on average enough of the drug to induce return. One possible explanation is that apparently independent sellers are retail agents for a single organization. The organization has an incentive to encourage users to return to that location; locations rather than sellers develop reputations. There is evidence of localized territoriality, perhaps not enough to generate market power (given the ease with which buyers can test other locations) but sufficient to induce competition through predictability.

That, however, is a model to be tested, not an assumption to be made. Organizational forms evolve over the epidemic, again perhaps because of variations in enforcement risk and dealer characteristics. In the mid-1980s, when the market for crack cocaine first appeared in many cities in the United States, it was characterized by three factors: *(a)* relatively modest law enforcement intensity; *(b)* new users who perhaps were not well known to each other; and *(c)* a low rate of dependence among primarily young users. By the mid-1990s, everything had changed: the probability of incarceration had increased substantially, the buyers and sellers had formed a stable group (at least over a period of, say, two years, which was long enough to allow for the average incarceration) and a large portion of purchases were made by dependent users who had been in the market for at least 10 years. Under such circumstances, the optimal price and purity strategy for dealers may vary, but it is not obvious how. The greater "churning" in the later stages of the epidemic creates incentives for cheating, but the loss of seller anonymity in a market of experienced participants counteracts that. Modelling such a phenomenon is likely to require use of dynamic game theory.

Conclusions

Drugs are sold in markets. The prices are determined systematically: notwithstanding the high rate of observed variation, they have clear patterns. Moreover, those prices have important implications both for participants and for others, including potential users (because prices affect the incentive to start using drugs) and society more generally (through crime and the generation of criminal income). Understanding how the price system functions should help to inform those who make drug policy decisions in a number of areas, including in the allocation of control resources between law enforcement and treatment and prevention and in the allocation of law enforcement resources for targeting buyers and sellers.

Markets for some illegal drugs appear to differ from conventional markets in fundamental ways. While the comparative statics methods developed for markets generally may carry over to those drug markets, the dynamic behaviour of the latter may require new tools.

Analysis of drug law enforcement and drug markets has been something of an intellectual desert, that is, a large territory with few occupants in widely scattered settlements. The body of theoretical literature is limited and is detached from any empirical work. There are only a few modelling efforts with an empirical base [21]. Drug markets are more difficult to model than conventional markets because data on drugs are more difficult to obtain; however, more sophisticated modelling of the kind being undertaken by some researchers, namely, Tragler, Feichtinger, Caulkins and Behrens [6, 7], is an important step in the right direction.

References

1. P. Reuter, *Disorganized Crime: the Economics of the Visible Hand* (Cambridge, MIT Press, 1983).
2. P. Reuter and M. Kleiman, "Risks and prices: an economic analysis of drug enforcement", *Crime and Justice: an Annual Review,* M. Tonry and N. Morris, eds., vol. 9 (Chicago, University of Chicago Press, 1986), pp. 128-179.
3. C. Peter Rydell and Susan S. Everingham, *Controlling Cocaine: Supply vs. Demand Programs* (Santa Monica, California, RAND, 1994).
4. G. Becker and K. Murphy, "A theory of rational addiction", *Journal of Political Economy,* 1988.
5. Susan Everingham and C. Peter Rydell, *Modeling the Demand for Cocaine* (Santa Monica, California, RAND, 1994).
6. D. A. Behrens and others, "A dynamic model of drug initiation: implications for treatment and drug control", *Mathematical Biosciences,* vol. 159, 1999, pp. 1-20.
7. D. A. Behrens and others, "Optimal control of drug epidemics: prevent and treat—but not at the same time?", *Management Science,* vol. 46, No. 3 (2000), pp. 333–347.
8. Peter Reuter, Robert J. MacCoun and Patrick J. Murphy, *Money from Crime: the Economics of Drug Selling in Washington, D.C.* (Santa Monica, California, RAND, 1990).
9. P. Reuter, P. Ebener and D. McCaffrey, "Patterns in drug use", in *When Drug Addicts Have Children: Reorienting Society's Response,* D. Besharov, ed. (Washington, D.C., American Enterprise Press, 1994), pp. 3-31.
10. P. Reuter, "On the consequences of toughness", *Searching for Alternatives: Drug Control Policy in the United States,* Edward Lazear and Melvyn Krauss, eds. (Stanford, California, Hoover Institution Press, 1991), pp. 138-162.
11. M. Kleiman, "Enforcement swamping", *Computer and Mathematical Modeling,* vol. 17, No. 2 (1993).
12. J. Cave and P. Reuter, *The Interdictor's Lot: a Model of Drug Interdiction* (Santa Monica, California, RAND, 1988).
13. J. Caulkins and P. Reuter, "What can we learn from drug prices?", *Journal of Drug Issues,* vol. 28, No. 3 (1998), pp. 593-612.

14. R. Needle, *Expenditure Patterns of Out-of-Treatment Drug Users* (Rockville, Maryland, National Institute on Drug Abuse, 1994).

15. Abt Associates, Illicit Drugs Price/Purity Report (Washington, D.C., United States Government, Office of National Drug Control Policy, 1998).

16. Don Weatherburn and Bromwyn Lind, "The impact of law enforcement activity on a heroin market", *Addiction,* vol. 92, No. 5 (1996).

17. David Simon and Edward Burns, *The Corner* (New York, Broadway Books, 1997).

18. R. Axelrod, *The Evolution of Cooperation* (New York, Basic Books, 1989).

19. A. M. Rocheleau and D. Boyum, "Heroin users in New York, Chicago, and San Diego", Washington, D.C., United States Government, Office of National Drug Control Policy, 1994.

20. K. J. Riley, *Crack, Powder Cocaine, and Heroin: Drug Purchase and Use Patterns in Six U.S. Cities* (Washington, D.C., United States Department of Justice, National Institute of Justice and Office of National Drug Control Policy, 1997).

21. J. Caulkins and others, *Mandatory Minimum Drug Sentences: Throwing Away the Key or the Taxpayers' Money?* (Santa Monica, California, RAND, 1997).

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

The dynamic character of drug problems*

J. P. CAULKINS

H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Drug Policy Research Center, RAND Corporation, Pittsburgh, United States of America

ABSTRACT

The author of the present article makes three points. Firstly, drug-related measures, such as the number of users, have changed rapidly over time, suggesting that they are not merely symptoms of underlying trends in the economy, demographics or other aggregates that change more slowly over time. Secondly, drug markets are subject to a wide range of feedback effects that can induce non-linearity into dynamic behaviour. Thirdly, there are at least five classes of drug epidemic models that reflect such non-linear dynamic behaviour. Some of those classes tend to be optimistic about the ability of drug control interventions to reduce use; others are pessimistic. It is hoped that the present article and, in particular, the typology, will inform and elevate the debate about drug policy, though it is unlikely to resolve that debate because of the inability to demonstrate empirically which classes of model is (are) more accurate.

Introduction

The thesis of the present article is that drug problems are dynamic phenomena characterized by non-linearity and feedback. To the extent that this is true, it is important to analyse drug problems with tools that recognize and handle that complexity. Regrettably, most of the literature on drug problems and policies applies linear, static and/or imprecise models.

The present article contains three sections. The first section examines empirical evidence for the thesis that drug problems are dynamic. The next section lists some of the principal sources of non-linearity and feedback in drug systems. The final section offers a typology of five models of drug epidemic. Which type is considered by a person to be most accurate can be an important determinant of that person's view about the potential role for policy.

*The research presented in this article was financed in part by the Austrian Science Foundation (FWF), the National Consortium on Violence Research and the National Science Foundation of the United States of America.

Evidence that drug problems evolve over time

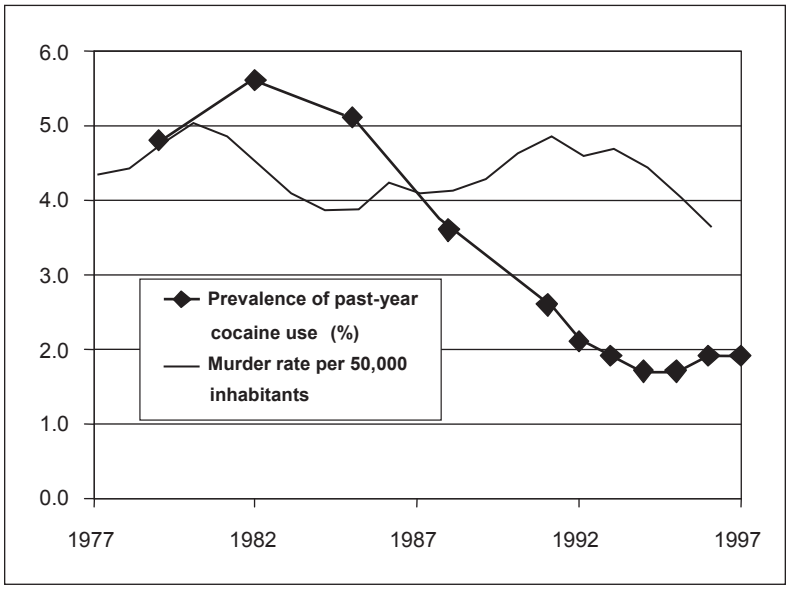
It is uncontroversial and uninteresting merely to assert that drug use and drug problems evolve over time. A stronger and more interesting argument is that they change more quickly and in more fundamental ways than do most other social phenomena. Such arguments cannot be quantified precisely, but it is clear that the variation in drug-related variables over the last 25-30 years has been quite substantial. Figures I-V illustrate this point by comparing variation in drug indicators (mostly pertaining to cocaine in the United States of America) with variation in series that are conventionally viewed as having been far from stable.

Violence is often described as having an epidemic component and the United States has witnessed sharp changes in the level of violence. Yet figure I shows that variation in the best-measured indicator of violence (the homicide rate) is much smaller than the variation in the number of Americans who are using cocaine (self-reported past-year prevalence as measured by the National Household Survey on Drug Abuse).

Likewise, much is made of the baby boom and baby boom echo. Rapid changes in the number of youth have stressed social institutions, have been blamed for variation in crime rates and are correlated with rates of youthful drug use. Yet figure II shows that the magnitude of those variations pale in comparison with the variation in the rate of cocaine use by youth (past-year and lifetime prevalence as assessed by the Monitoring the Future survey).

Initiation of illicit drugs also varies much more than does initiation of licit drugs. For instance, as estimated by the National Household Survey of Drug Abuse

Figure I. Variation in past-year cocaine use among the household population and homicide, 1977-1997



[1], between 1962 and 1992 the ratios of maximum to minimum numbers of initiates into regular alcohol and cigarette use were just a little over 2:1. For marijuana the ratio was about 15:1, and for cocaine it was almost 150:1 (see figure III.)

Figure II. Variation in youthful cocaine use exceeds variation in number of youth, 1975-1995

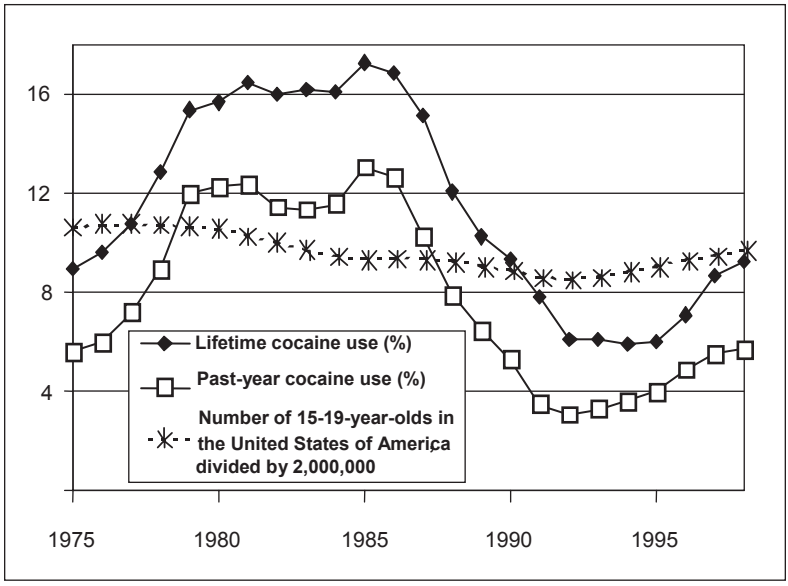
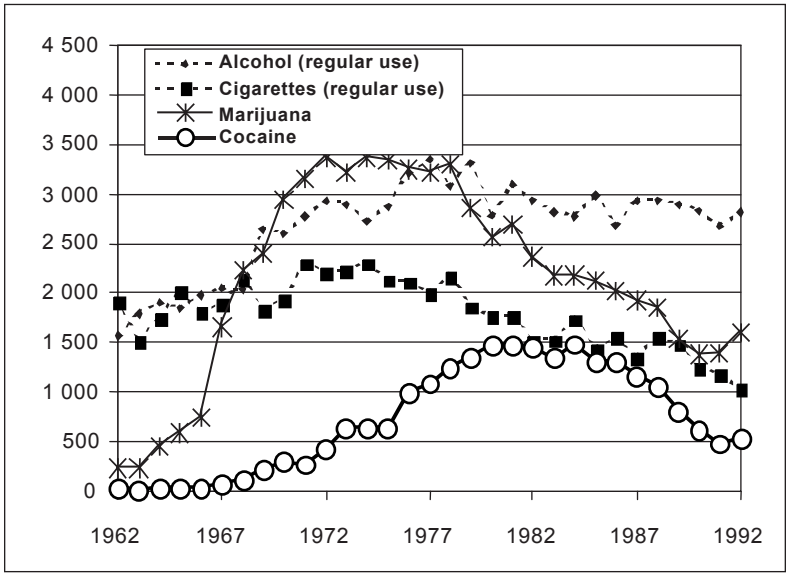


Figure III. Trends in incidence of cocaine, marijuana and regular cigarette use in the United States of America (thousands/year), 1962-1992



Drug prices have also been unstable. Among licit goods, oil prices are notoriously unstable, being driven to a sharp peak during the oil crisis associated with the fall of the Shah of Iran and falling sharply in subsequent years. Yet the decline in retail cocaine prices was every bit as steep (figure IV).

And the consequences of drug use have grown sharply over time. It is well known that the rate of new acquired immunodeficiency syndrome (AIDS) cases has grown from nearly zero before 1980 to epidemic proportions. The growth in the number of emergency room mentions for cocaine has been comparably swift (figure V).

These figures are meant to convey one simple point. Drug epidemics can, and the cocaine epidemic in the United States did, undergo very rapid change over time. None of the axes in the figures are false axes drawn to exaggerate the magnitude of modest changes. More such graphs could be drawn. Everingham and Rydell [2], for example, produced an often-cited chart showing how the mix of light and heavy cocaine users changed dramatically over time. In 1980 light users (those using cocaine less than once monthly) were responsible for about 60 per cent of cocaine demand in the United States; by 1990 that proportion had fallen to 30 per cent. In many ways, the cocaine problem of the 1980s was not the same as that of the 1970s, and the cocaine problem of the 1990s was not the same as that of the 1980s.

Examining a much larger period in history, Musto [3] notes alternating periods of greater and lesser drug use. In particular, a cycle of quiescence, rapid escalation, plateau and gradual decline has been observed for a number of drugs, including powder cocaine in the late nineteenth and early twentieth centuries [4] and crack in more recent times [5].

Figure IV. Declines in the price of cocaine and crude oil, 1978-1996 (1996 United States dollars)

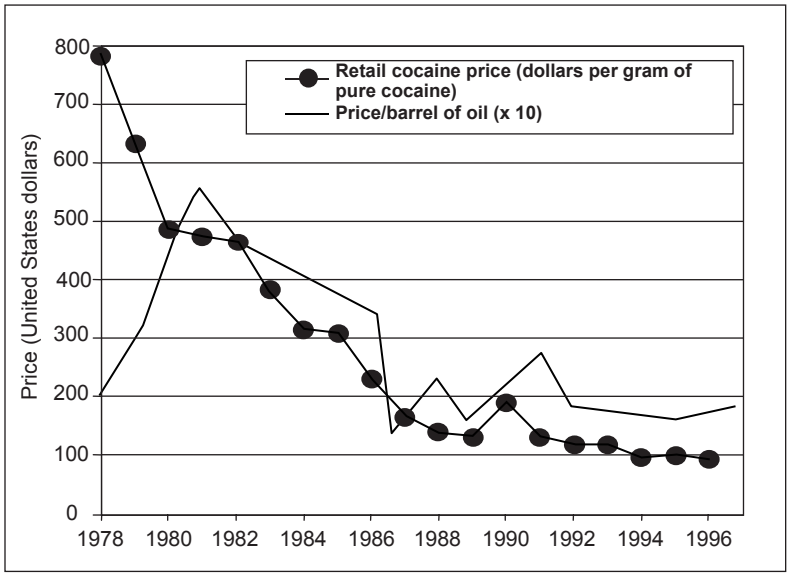
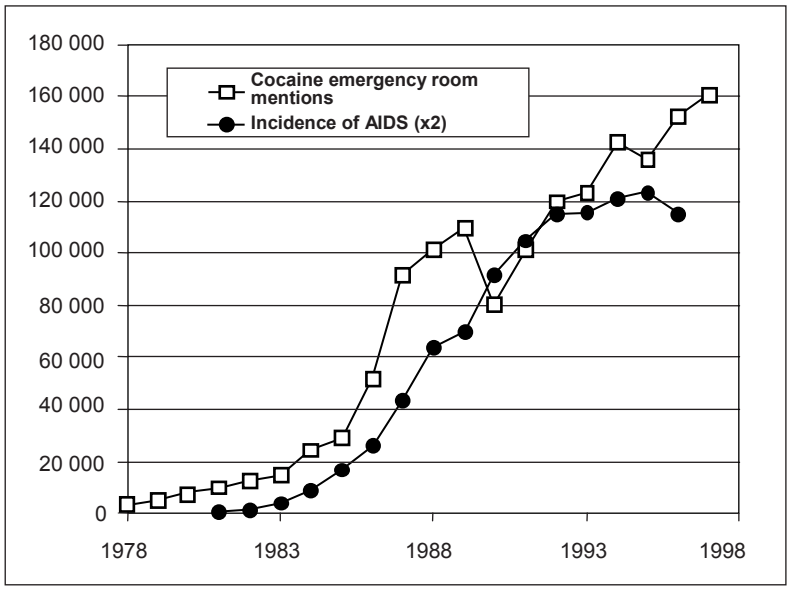


Figure V. Growth in emergency room mentions for cocaine and incidence of AIDS, 1978-1998



Some systems vary over time, primarily in response to variation in some exogenous forcing function. For example, the number of flowers in temperate regions varies around the year because of seasonal variation in temperature and amount of daylight. Other systems generate variation over time because of the character of their internal structure. For example, predator-prey models can generate cycles in levels of both the predator and the prey populations because of internal dynamics. When prey are plentiful, predator populations grow until they drive down the prey population. That can lead to starvation for predators, which allows the prey population to recover and the cycle to repeat itself.

An important question is whether cycles of drug use are driven primarily by exogenous factors (e.g. the business cycle) or internal structures (as with predator-prey models). The fact that drug-related phenomena vary so much more than do many other phenomena hints that there may be interesting internal dynamics to drug markets and patterns of drug use. What some of those dynamics may be is explored in the next section.

Sources of non-linearity and feedback in drug systems

Drug-related phenomena seem capable of changing much faster than underlying societal characteristics such as economic well-being, demographic variables or other health-related behaviours. Systems that change quickly often do so because they contain some feedback or non-linearity. Some likely sources of feedback or non-linearity in drug systems are identified below.

Enforcement swamping

Drug market participants, like people generally, respond to incentives. One important incentive is the risk of enforcement. That risk, in rough terms, is determined by the amount of effort expended by law enforcement agencies relative to the size of the market. For example, compare a small city that arrests 100 of its 500 drug sellers per year with a much larger city that has four times as many sellers (2,000) but makes only twice as many arrests (200) per year. The level of law enforcement activity is higher in the second city (200 arrests vs 100), but the enforcement pressure or intensity is greater in the first because 100 out of 500 is a greater proportion than 200 out of 2,000. In some sense individual market participants do not care how many people are arrested. They care selfishly only about their individual probability of arrest.

Responses to law enforcement pressure include reducing the frequency of offending and displacing the activity to another location, drug, or time of day [6]. In either case, increased law enforcement pressure can reduce the number of offenders who are subject to that pressure.

Together these two observations are sufficient to create a powerful feedback effect, which has been dubbed "enforcement swamping" [7]. Suppose the number of drug market participants increases for some exogenous reason, such as a shift in tastes. The expanded market dilutes the given level of enforcement over a larger denominator, reducing the enforcement pressure experienced by any given participant. That reduced enforcement risk makes it more appealing for others to join the market, which further dilutes enforcement pressure. Depending on the specific circumstances, the feedback effect could grow out of control (possibly "tipping" the market to a new, higher-level equilibrium) or it could merely amplify the effect of the original exogenous change.

The same feedback effect can operate in reverse. Suppose the authorities decide to increase the number of arrests. That increases the intensity of arrests, which might induce some people to cease or relocate their drug activities. If so, the resulting reduction in market size further increases the enforcement pressure borne by those who remain, which might in turn encourage still others to exit. Again, this feedback might push the market to a new type of equilibrium (e.g. eliminating the market altogether) or it might merely amplify the effect of the original change in enforcement level, but in either event represents a non-linearity.

Individual demand is a function of past levels of consumption

Economists are careful to distinguish two related concepts: demand and consumption. The technical definition of consumption is the same as the lay definition. It refers to the amount of a good produced, sold and consumed in a market. Demand is different. Demand is not a single quantity but a relationship between price and consumption. It describes how much consumers would want to purchase as a function of price. Typically, consumers would purchase more of a good if prices were low than if they were high. This relationship between the market price and the quantity consumed is often drawn on a graph and referred to as a demand curve.

For many goods, the quantity consumed varies over time with market conditions but the demand curve is stable, or, if it varies, it varies because of exogenous factors. The demand for luxury goods, for example, may be higher during strong economic periods.

For drugs, demand is not fixed. It is a function of past consumption because of addiction and tolerance. The economic interpretation of addiction is subtle and still evolving [8], but one interpretation is that past consumption increases the value of future drug consumption relative to the value of consumption of other goods. This manifests itself in various ways, including the observation that, as some people become addicted, they spend a larger and larger share of their disposable income on the drug. Thus, drug consumption is reinforcing in an economic as well as a psychological sense.

Tolerance can have a similarly reinforcing effect if users seek out ever-increasing doses to achieve the same effect. It can also have the opposite effect if it reduces the psychic effect or benefit of a given amount of consumption. Which effect dominates depends on a variety of circumstances, including the type of drug. Anecdotal evidence suggests that tolerance reinforces future consumption for heroin but constrains it for Ecstasy (methylenedioxyamphetamine (MDMA)).

Addiction and the positively reinforcing effects of tolerance can create a positive feedback effect. Suppose supply increases. That has no immediate effect on demand, but would increase consumption. For a conventional good, that would be the end of the story; however, for drugs, that increase in consumption can subsequently lead to an increase in demand, which increases consumption still further, which increases demand, and so on. Whether that positive feedback pushes the market to some qualitatively different equilibrium or merely amplifies the effect of the original shift in supply depends on the particular circumstances, but in either case represents a non-linearity.

These demand-amplifying effects are not unique to drugs. Goods that are an "acquired taste" (opera is a common example) have a similar character and network externalities can make demand a function of past consumption. For example, demand for electronic mail (e-mail) grew as subscriptions to e-mail did because the value of e-mail depends in part on how many other people use it. Nevertheless, the fact that there are other exceptions to the "standard" notion of stable demand does not undermine the importance of this feedback for drugs.

Initiation is a function of current levels of use

Drug use is often described as being "contagious". The metaphor is appropriate, even though there is not a physical, pathogenic infection vector as with malaria because initiation rates are significantly influenced by the current prevalence, or level, of use. In particular, most people who start using drugs do so through contact with a friend or sibling who is already using them. Indeed, the metaphor of a drug "epidemic" is commonly used precisely because of that tendency for current users to "recruit" new users.

The feedback from current use to initiation is not necessarily uniformly positive. Musto [3] has argued that, in addition, knowledge of the possible adverse effects of drug use acts as a deterrent or brake on initiation. He hypothesizes that drug epidemics eventually die out when a new generation of potential users becomes aware of the dangers of drug abuse and, as a result, does not start to use drugs. Whereas many light users work, carry family responsibilities and generally do not manifest obvious adverse effects of drug use, a significant fraction of heavy users are visible reminders of the dangers of using addictive substances. Hence, large numbers of heavy users might be expected to suppress rates of initiation into drug use.

Learning by doing

Economists also distinguish between the quantity supplied in a market (which is the same as the quantity consumed) and the supply curve. The supply curve, like the demand curve, is not a single quantity, but a schedule or relation that describes how much suppliers would be willing to sell as a function of the market price. Again, as with demand, for the typical good, the supply curve is usually thought of as stable or as varying only in response to exogenous factors; however, for drugs, the supply curve can itself be a function of past production because of what Kleiman calls "learning by doing" [9].

"Learning by doing" refers to the idea that the supply curve is directly affected by the cost of production, and production costs decline as suppliers get more experience. The more a supplier organization has produced, the more chances it has had to discover more efficient means of production.

Again, this type of feedback is not unique to drugs. It occurs with many emerging industries; for example, prices for the electronic calculator collapsed as production volumes led to innovation. Even though drug use has occurred for millennia, the existing illicit drug markets are relatively new. High-volume cocaine production is less than 30 years old.

This phenomenon is more pervasive for illicit drug markets, however, because enforcement generates a constant turnover among drug suppliers and supply tactics, so at any given time many individual suppliers may be working up a learning curve even as the industry as a whole matures. Incapacitation drives some of that turnover. When experienced suppliers are incarcerated they are replaced by novice sellers. Avoidance plays a role as well. When improvements in law enforcement force suppliers to change smuggling routes or tactics, suppliers start over on a new learning curve for that route or tactic.

Such effects can operate at the market level as well as the organizational level. If interdiction forces smugglers to use a new trans-shipment country, initially smuggling may be costly, but if law enforcement agencies in the new trans-shipment country become corrupt over time, smuggling costs may decline.

The potential for positive feedback loops with "learning-by-doing" effects is clear. The more that is sold, the more efficient suppliers become. The more efficient suppliers become, the lower prices will be, and lower prices induce greater consumption, which leads to further "learning by doing", and so on.

Types of epidemic models

These and other feedback effects permit an almost infinite number of models to be created. Classifying them by mathematical structure (e.g. discrete vs continuous time models) is of limited value, but five classes of model may be identified on the basis of their explanation of the one empirical regularity concerning drug use about which there is little debate. Levels of use have been observed to rise rapidly from relatively low levels to much higher levels.

As yet, it is less clear what happens to drug use after that rise. Some evidence suggests that drug use remains at the new higher levels for an extended period of time; for example, the number of heroin addicts in the United States does not seem to have ebbed significantly after its rapid increase in the late 1960s and early 1970s. Some evidence suggests that use falls off from its peak but never returns to its original, low level; for example, marijuana use in the United States is well below peak levels, but remains far above levels of the period preceding the 1960s. And there is some evidence that the level of drug use can return to quite low levels, at least for a time; for example, the peak in cocaine use in the United States at the beginning of the twentieth century was separated from the current cocaine epidemic by a period (1930-1965) during which cocaine use was much less common.

The ambiguity about what happens after an explosion in drug use means that several types of epidemic models are consistent with the minimal facts about which there is clear consensus—namely, that drug use can rise rapidly from low to high levels. In particular, there are at least five broad classes of model of drug use.

The first class of model assumes that control drives everything. The internal dynamics of the drug epidemic play at most a secondary role. To those who subscribe to that view, if drug use is low, it is because drug policy is successful. Conversely, if drug use is high, that is clear evidence of a failure of policy. An explosion in drug use can easily be explained as a precipitous decline in the effectiveness of control efforts. Perhaps understandably, much of the debate among policy makers implicitly if not explicitly adopts this view that policy is central. It encourages evaluating control efforts with simple "before-and-after" comparisons. Counterfactuals are irrelevant.

A variation on this class of model assumes that drug use is always threatening to grow exponentially without bound and the only thing preventing every person or at least every child from using drugs is the control efforts that are in place. This version is a convenient one for agency administrators to adopt when seeking to justify their budgets. It is also not refutable. There is an old joke about a person snapping his fingers who, when asked why, replies "To keep the elephants away". When informed that there are no elephants in the area, he triumphantly concludes that finger-snapping is an effective means of elephant control. Likewise, to those who subscribe to this model, the existence of non-users justifies continued funding of drug control efforts.

Given the foregoing discussion of drug epidemic dynamics and feedback, the reader can safely surmise that the author of the present article does not subscribe to this view.

The second class of model is as pessimistic about the power of drug control interventions as the first is optimistic. In it, the only stable level of use is a high level of use. Within the model, low levels of use are seen as unstable, transient periods. This transient character is reconciled with the observed persistence of low levels of drug use by invoking some exogenous shock. That is, the model is assumed to apply only after some exogenous structural change in conditions. Once that change occurs, the low levels of use are no longer sustainable and use explodes. For example, one might view high levels of cocaine use in the United States as inevitable if there is an efficient supply pipeline connecting the United States to source countries in South America. That supply line was tenuous before the 1970s, so use could remain at low levels. Once it was established, use rapidly expanded and, according to this view, there is little prospect for serious reduction in use without some other exogenous shock to the system (such as elimination of coca production in South America by some blight or fungus).

In this view of the world, routine drug control efforts are of little consequence once use has approached its high-level equilibrium. They might push use down a little, but unless some dramatic intervention manages to alter the structure of the system, control efforts will have only marginal effects.

The model of Tragler and colleagues [10] is of this character, with the exception that when drug use is low, control can suppress it. Thus, when use is very low, the model is like the first class of model. There would be explosive growth, but effective control prevents it. However, when use is high, "enforcement swamping" vitiates the power of controls.

In the third class of model, initially there are no drug users but there are many "susceptibles" who are vulnerable to drug initiation if the opportunity arises. When the drug is first introduced to this population, drug use rapidly infects everyone who is susceptible; however, not everyone is susceptible, so not everyone becomes a drug user. Furthermore, people do not use drugs for ever and "susceptibles" who quit using drugs are no longer vulnerable to "infection". So, in the long run, initiation is restricted to people who are new arrivals in the system (e.g. new birth cohorts). Thus, an explosion in drug use can be seen when the drug is introduced to an unexposed population, but that explosion is followed by a decay to a lower endemic level as people mature out of drug use.

In these models (typified by Rossi [11]) the dimensions of the epidemic are determined by the proportion of the population that is susceptible and the typical duration of drug use. Everyone who is susceptible will get "infected" and continue to use drugs for however long people use them. Controls that operate on these two parameters are meaningful: prevention programmes that inoculate people against drug use or treatment programmes that shorten drug use careers can reduce the population of drug users. Other interventions, however, tend to have only modest effects, for example, slightly delaying the inevitable explosion in drug use.

The fourth class of model, so-called "tipping" models, suggests an intermediate degree of optimism about the role of policy. "Tipping" models are characterized by (at least) two stable equilibria, one at a low level of drug use and one with a high level of drug use. Either low or high levels of drug use can persist

indefinitely in the absence of some intervention or exogenous shock. These models view explosions in drug use as instances of "tipping" from the low- to the high-level equilibrium. Their implications for policy are twofold. Firstly, policy makers should do whatever they can to prevent the system from "tipping" from low to high levels of drug use. Typically that recommendation is of little value because the problem only attracts serious analysis after it has "tipped" into the high-level (problematic) state. Secondly, modest interventions are unlikely to have much effect, but a truly massive intervention might succeed in "tipping" the system back to a low-level equilibrium, at which point the level of intervention could be cut dramatically without having drug use return to its high levels. Hence, these models tend to suggest that it would be better to pursue a relatively modest control programme or to be very aggressive for a period long enough to "tip" the system back to its low-level equilibrium, at which point control can return to lower levels. Caulkins [12] and Baveja and colleagues [13, 14] offer examples of this type of model.

In the fifth class of model, drug use grows rapidly at first because of some positive feedback but, over time, negative feedback pushes drug use back down. Egan's journalistic description [15] of the ebbing of the crack epidemic in New York belongs to this class. Behrens and colleagues [16, 17] offer a more mathematical model in which drug use initially spreads exponentially, but prolonged drug use leads to adverse consequences that give the drug a negative reputation that suppresses initiation. By differentiating between light and heavy users, this model can endogenously create recurring epidemics of drug use separated by intervals of low drug use without invoking exogenous shocks or interventions to generate those cycles. Drug control interventions may or may not be highly valued in these models, depending on the details of the model and parameter values.

On the one hand this typology is very useful. When two people disagree fundamentally about the nature of drug epidemics or the efficacy of drug control interventions, it is often because their respective models (whether formal or intuitive) belong to different classes. Figuring out which classes people subscribe to can cut to the heart of the disagreement.

On the other hand, this typology is not very useful inasmuch as there is as yet no empirical means of validating one of these classes of model or disproving another. Model validation is tricky in general. In the drug policy domain one faces the added burden of a paucity of reliable data and an inability to run controlled experiments. So those people may simply have to agree to disagree about which class of model is most appropriate.

Discussion

Drug use and associated phenomena change rapidly over time. In particular they change much more rapidly than do most other macro-level measures of social conditions, suggesting that such dynamics are, to a significant degree, driven by endogenous and not exogenous factors. This empirical observation is complemented by theoretical and qualitative depictions of drug market dynamics.

Particulars of the dynamic evolution of drug use vary by substance, time and location, and data characterizing those changes are relatively weak. Nevertheless, one empirical regularity stands out: drug use can and all too frequently does rise very rapidly from quite low to quite high levels.

The present article defines five broad classes of drug epidemic model that are consistent with such rapid escalation in drug use. They vary sharply in their implications for the ability of drug control interventions to materially influence drug use. A person who subscribes to the second class of model (high levels of drug use are the only stable condition) may disagree strongly with a person who subscribes to the third class (all "susceptibles" have a high probability of becoming "infected") about whether a new drug epidemic will ebb of its own accord and will disagree with a person subscribing to any of the other classes of model about the benefits of drug control interventions.

Most people implicitly adopt one or another of these classes of model, but few consciously realize that they have done so. Bringing more explicit recognition of the models and their implications into discussions on drug policy may help resolve differences of opinion or at least concisely identify the sources of disagreement. In the longer run, a concerted effort to refine the models and collect data to support their validation and parameterization could elevate the precision and utility of drug policy analysis considerably.

References

1. Substance Abuse and Mental Health Services Administration, *National Household Survey on Drug Abuse: Main Findings* (Washington, D.C., United States Department of Health and Human Services, 1999).
2. S. S. Everingham and C. P. Rydell, *Modeling the Demand for Cocaine* (Santa Monica, California, RAND Corporation, 1994).
3. D. F. Musto, *The American Disease: Origins of Narcotic Control* (New York, Oxford University Press, 1987).
4. Joseph Spillane, "Did drug prohibition work? Reflections on the end of the first cocaine experience in the United States, 1910-45", *Journal of Drug Issues,* vol. 28, No. 2 (1998).
5. Andrew Lang Golub and Bruce D. Johnson, "Crack's decline: some surprises across U.S. cities", *Research in Brief* (Washington, D.C., National Institute of Justice, 1997).
6. Jonathan P. Caulkins, "Thinking about displacement in drug markets: why observing change of venue isn't enough", *Journal of Drug Issues,* vol. 22, No. 1 (1992), pp.17-30.
7. Mark A. R. Kleiman, "Enforcement swamping: a positive-feedback mechanism in rates of illicit activity", *Mathematical and Computer Modeling,* vol. 17, 1993, pp. 65-75.
8. Gary S. Becker and Kevin M. Murphy, "A theory of rational addiction", *Journal of Political Economy,* vol. 96, No. 4 (1988), pp. 675-700.
9. Mark A. R. Kleiman, *Marijuana: Costs of Abuse, Costs of Control* (Westport, Connecticut, Greenwood Press, 1989).

10. G. Tragler, J. P. Caulkins and G. Feichtinger, "The impact of enforcement and treatment on illicit drug consumption", Working Paper 212, Department of Operations Research and Systems Theory, Vienna University of Technology, 1997.
11. Carla Rossi, "A mover-stayer type of model for problematic drug use epidemics", *Bulletin on Narcotics* (United Nations publication), vol. 53, Nos. 1 and 2 (2001).
12. Jonathan P. Caulkins, "Local drug markets' response to focused price enforcement", *Operations Research,* vol. 41, No. 5 (1993), pp. 848-863.
13. Alok Baveja and others, "Modeling the response of illicit drug markets to local enforcement", *Socio-Economic Planning Sciences,* vol. 27, No. 2 (1993), pp. 73-89.
14. Alok Baveja and others, "When haste makes sense: cracking down on street markets for illicit drugs", *Socio-Economic Planning Sciences,* vol. 31, No. 4 (1997), pp. 293-306.
15. Timothy Egan, "A drug ran its course, then hid with its users", *New York Times,* 18 September 1999.
16. Doris A. Behrens and others, "A dynamical model of drug initiation: implications for treatment and drug control", *Mathematical BioSciences,* vol. 159, 1999, pp. 1-20.
17. Doris A. Behrens and others, "Optimal control of drug epidemics: prevent and treat—but not at the same time", *Management Science,* vol. 46, 2000, pp. 333-347.

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

Key variables and data requirements in modelling drug systems

E. SINGLE

Professor of Public Health Sciences, University of Toronto, and Senior Research Associate, Canadian Centre on Substance Abuse, Canada

Abstract

The author of the present article explores factors that account for variation in drug use and drug-related harm over time. A model of a drug system is presented, consisting of dependent variables, proximate causes and underlying determinants. The dependent variables concern drug use (level and pattern), adverse health consequences and drug-related crime. The major proximate causes of drug use and drug-related harm include drug availability (price and physical availability), attitudes towards use (fear of legal sanctions, health beliefs regarding risks and cultural beliefs) and alternatives to the illicit market (home cultivation and legal intoxicants for users, alternative career and income prospects for sellers). The underlying determinants that influence those proximate causes include policy (drug laws, preventive education and risk management programming) and environmental factors (geographical isolation, climate and fauna, and threat of acquired immunodeficiency syndrome (AIDS) and other infectious diseases). Despite recent improvements in the measurement of problematic patterns of drug use, there is still a paucity of data on patterns of use and drug-related harm. The viability of drug policy is thus often measured in terms of changes in levels of drug use and/or changes in the number of persons detected and charged with drug crimes. Until valid and reliable data are available on its key variables, any model of drug systems will have very limited applicability.

Introduction: characteristics of systems of drug information

The author of the present paper explores aspects of drug systems that may account for a substantial portion of the variation in drug use and drug-related harm over time in a particular setting. The utility of any model depends on the availability of data on key indicators. Before focusing on the key variables that have an impact on drug use and drug-related problems, the discussion begins with three observations concerning the characteristics of drug information systems.

The limits of drug information systems

Firstly, there is a wide variation in drug problems and responses to those problems. Therefore, even though uniformity in drug information systems can and should be promoted, there probably can never be a perfect information system for all countries. The nature of drug problems differs in different settings [1]. In regions such as South America and South-East Asia where a large share of the world's cocaine and heroin are produced, the illicit drug trade has created substantial underground economies with consequent problems of law enforcement and economic control. In the consuming countries in Europe and North America, the problems tend to be the adverse health, social and economic consequences of the use of illicit drugs.

Furthermore, there is also a wide variation in national drug policies. Responses to illicit drug problems range from strict enforcement of punitive drug laws to benign neglect. In parts of Australia, Europe and North America harm reduction policies have been implemented to reduce the adverse consequences of illicit drug users who cannot be expected to cease their drug use at the present time [1]. Even within countries, there are often cycles of panic over emerging drug problems, followed by periods of indifference when other pressing issues push illicit drugs to a relatively low place on the national policy agenda [2].

Although international treaties provide a common framework for drug policy and a certain degree of uniformity in social responses to drug problems, there is inevitably considerable variation in the nature and magnitude of illicit drug problems as well as the social responses to those problems, both between countries and regions and even within countries over time.

Data systems depend on well-articulated goals and performance indicators

The second observation is that without a clearly articulated set of goals and performance indicators, drug information systems will inevitably fail to meet their objectives. This may seem obvious, but the fact remains that the goals and performance indicators for national drug strategies are rarely well articulated. Focus is often placed on reducing the prevalence of any illicit drug use rather than specific indicators of drug-related harm. Examples of reasonably well-articulated performance indicators for national drug strategies are relatively uncommon. For example, Australia's National Drug Strategy has been subject to evaluation [3] and its goals and performance indicators have been articulated in a National Drug Strategic Plan [4].

Canada's Drug Strategy currently suffers from a lack of dedicated funding, but its goals have been articulated as follows [5]:

(a) To reduce the demand for drugs;

(b) To reduce substance-related mortality and morbidity;

(c) To improve the effectiveness of and accessibility to information, education and other interventions, including treatment and rehabilitation;

(d) To restrict the supply of illicit drugs.

Each of those goals is further classified into sub-objectives with associated performance indicators. For example, performance indicators for the goal of reducing demand for drugs include not only reductions in levels of drug use, but also increases in attitudes that inhibit drug use (e.g. realistic health concerns), reductions in patterns of illicit drug use that are particularly likely to result in harm (e.g. regular use, injection drug use, needle-sharing) and particular targets for reductions in those indicators by high-risk groups such as native peoples, street youth, prisoners and other socially disadvantaged groups [5]. Similar sub-objectives and performance indicators are articulated for the other goals of the national strategy. The specific goals and performance indicators will vary between countries, as they should, because, as noted above, the nature of problems caused by illicit drugs varies between countries.

Thus, the task for internationally comparable drug information systems is to find a common set of goals and performance indicators that national drug policies can build upon and expand within their own national context. Data sources should also be identified, in order to determine information gaps and set research agendas. The key point is that the articulation of goals and specific performance indicators is required to create the framework for drug information systems that ultimately will provide the data to test any dynamic model of drug use and drug problems. Where the goals and performance indicators of a national drug strategy are not clearly articulated, drug information systems will be based on the availability of data rather than data requirements. The development of a dynamic drug model based on poorly articulated data systems may be doomed to failure.

Greater attention is required on economic aspects of illicit drug abuse

The third observation concerning drug information systems is that there is a strong need for more rigorous and comprehensive economic data on substance abuse to promote evidence-based decision-making and a more consistent response to substance abuse. The economic ramifications of illicit drugs are not well understood, either in the producing countries or the consuming countries and regions.

Four key questions need to be addressed to help policy makers to make well-informed decisions on drug issues [6]:

What is the cost of drug abuse to society?

What portion of those costs are realistically avoidable?

What and where should policy makers invest to avoid those costs?

How well are such investments performing over time?

Researchers have given substantial attention to the costs of alcohol and tobacco use [7]. In contrast, relatively little attention has been given to the costs

that society bears as a result of the abuse of illicit drugs. That deficiency can be attributed largely to data problems inherent in any attempt to quantify the social costs of illicit drugs. It is difficult to quantify the production, consumption, import, export or price of illicit drugs. In addition, although significant information is available on the causal links between drug abuse and health, the causal links in other areas, crime in particular, are extremely difficult to quantify. For these reasons little quantitative information exists on the social costs of illicit drug abuse.

A proposed dynamic model for drug systems

The first step in modelling the dynamic aspects of drug systems is to identify the key variables and data requirements in a model that could explain variations in drug use and drug-related problems in a particular setting over time. It is proposed that a model of a drug system consists of dependent variables, proximate causes and underlying determinants:

(*a*) Dependent variables:
- (i) Drug use variables:
  - a. Levels of drug use;
  - b. Patterns of use: heavy, dependent use;
  - c. Injection drug use;
  - d. Associated risk behaviours such as needle-sharing;
- (ii) Drug-attributable crime:
  - a. Drug violations;
  - b. Property crime attributable to drug use;
  - c. Violent crime attributable to drug use;
- (iii) Other adverse health and social consequences:
  - a. Overdose, suicide;
  - b. Other health effects;
  - c. Abuse, family discord;
  - d. Poor school or work productivity;

(*b*) Proximate causes:
- (i) Drug availability;
- (ii) Attitudes towards illicit drug use;
- (iii) Alternatives to using or dealing in illicit drugs;

(*c*) Underlying determinants:
- (i) Environmental factors;
- (ii) Drug policy:
  - a. Interventions with users:
    - i. Risk reduction among continuing users;
    - ii. Treatment and rehabilitation;
    - iii. Social welfare policy bearing on users;
  - b. Demand reduction;
  - c. Supply reduction.

The major proximate causes of drug use and drug-related harm include, but are not restricted to drug availability (price and physical availability), attitudes towards use (fear of legal sanctions, health beliefs regarding risks and cultural beliefs) and alternatives to the illicit market (home cultivation and legal intoxicants for users, alternative career and income prospects for sellers). The underlying factors that may influence those proximate causes include, but are not restricted to, policy (drug laws, preventive education and harm reduction programming) and environmental factors (geographical isolation, climate and fauna, threat of acquired immunodeficiency syndrome (AIDS) and other infectious diseases).

Dependent variables in the model

As mentioned above, there are three main dependent variables in the model: levels and patterns of drug use; adverse health and social consequences; and drug-related crime. Information is needed not only on the proportion of the population using illicit drugs, but also which drugs, in what combinations, by what mode of administration, in which settings and whether the drugs are used in a manner that minimizes or maximizes the chances that serious health and social consequences will occur. The key point concerning drug use variables is that a robust model requires more than simply information on levels of drug use. Without detailed information on patterns of use, relatively little variance in drug-related problems will be explained by the model.

Illicit drug use is associated with a variety of problems, including crime, afamily dysfunction, workplace problems and health disorders [8]. Illicit drugs are among the leading causes of preventable death and illness among young persons and the social and economic costs related to illicit drug use are considerable. The problems of illicit drug use negatively affect many communities, making neighbourhoods unsafe, diminishing property values and diverting limited police resources from other pressing needs. The major health problems associated with illicit drug abuse are suicide, drug overdose and communicable diseases such as AIDS and hepatitis C. Other adverse health and social consequences of drug use include hospitalization or other treatment for drug dependency, lower economic productivity, poor school performance, child and spousal abuse and family discord.

Drug use is related to crime but the degree to which the relationship is causal is unclear. Chronic or dependent use of the so-called "hard" drugs—heroin, cocaine or crack, speed, lysergic acid diethylamide (LSD) and other strong hallucinogens—is often implicated as a contributory cause of property crime, in particular burglary and theft. Assault, homicide and other crimes of violence have resulted from "turf wars" in the illicit drug market. Illicit drug users are disproportionately involved in incidents of spousal and child abuse. Even cannabis use has been implicated as a contributory cause of crime, namely, the crime of impaired driving [9]. Criminal offenders have disproportionately high rates of illicit drug use [10,

11]. Up to 80 per cent of Canadian criminal offenders reported using illicit drugs during their lifetime, 50-75 per cent showed traces of drugs in their urine at the time of arrest and close to 30 per cent were under the influence of drugs when they committed the crime for which they were accused [12]. Drug addicts admitted to treatment often have criminal records [13].

There is clearly a strong relationship between illicit drug use and crime, but the fact that a crime is committed by someone using illicit drugs does not necessarily mean that the drug use caused the crime to be committed. The pharmacological effects of drugs and the need for addicts to commit crime to support their drug habit are at best only partial explanations for the link between drug use and crime. The majority of illicit drug users are not dependent and most users, even dependent users, do not commit property crimes [14]. Most addicts who commit crimes began doing so prior to becoming drug-dependent [15]. Addicts who commit property crimes tend to use drugs at very high levels, they have few legitimate sources of income and in the majority of cases they were engaging in criminal behaviour prior to drug use [12]. Furthermore, many former addicts continue to commit property crimes after they no longer use drugs [16].

A more plausible explanation for the strong association between illicit drug use and crime is that addicts adopt a deviant way of life that accounts for both their drug use and their criminal behaviour. A number of longitudinal studies have shown that drug use and criminality are related to a similar set of socio-demographic and personality variables, for example, poverty, poor future career or income prospects and low investment in social values [17-19]. There are undoubtedly many commonalities in the aetiology of both criminality and illicit drug use. Drug use and crime may well be mutually reinforcing, but according to that viewpoint, the real cause of both drug use and criminal behaviour is a complex set of underlying personality and social determinants.

Thus, the attribution of crime to illicit drug use is fraught with methodological difficulties and there is a lack of research on the proportion of all crime that can be causally attributed to illicit drug use [12, 20]. While there is some doubt concerning the extent to which drug use leads to crime, there is little doubt that crime results from systemic violence inherent in the illicit drug trade. Many crimes result from "turf wars" between rival distributors as well as arguments and robberies involving buyers and sellers on the illicit market [21]. Systemic violence in the illicit drug market is most common in economically and socially disadvantaged areas that have traditionally high rates of violence. It should also be noted that not only are drug addicts more likely to commit crimes, but they are also more likely to be victims of violent crimes [15].

Proximate causes of drug use and drug-related harm

The major proximate causes of drug use and drug-related harm consist mainly of drug availability, attitudes towards use and alternatives to the illicit market. Drug availability could be measured by survey questions regarding the ease with which respondents may obtain various illicit drugs or by tracking the street price per unit

of potency for illicit drugs. In fact, this is rarely done. In the absence of better information on drug availability, seizure data are often taken as a surrogate measure of availability. Indeed, large expenditure for drug trafficking enforcement is typically justified purely on the basis of the number of arrests and amount of drugs seized (sometimes accompanied by a large amount of cash as well as weapons). The ultimate impact on drug availability is rarely assessed. In what may be the only systematic study on the effects of drug trafficking enforcement, it was found that arresting drug traffickers tends to have local, short-term and negative impacts on drug problems [22]. Typically, major drug arrests were followed by a temporary increase in street prices for illicit drugs until new sources emerged and led to more property crime by dependent users [22]. Long-term impacts were negligible. While one study is hardly conclusive evidence, it would appear that law enforcement data on trafficking arrests and drug seizures are a poor measure of drug availability.

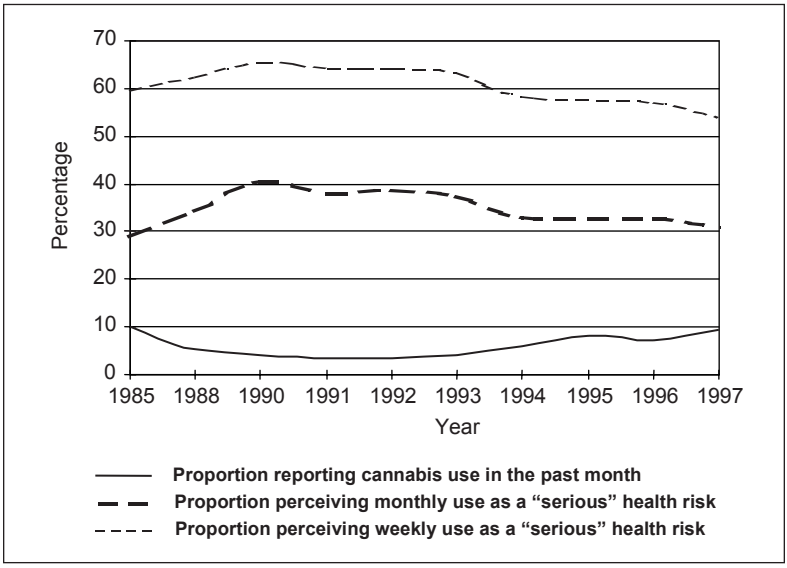
Attitudes towards drugs represent a key determinant of drug use. While attitudes towards heroin or cocaine use are rarely measured, it is reasonably well documented that changes in rates of cannabis use appear to be more strongly connected with changing perceptions of health risks rather than fear of legal sanctions or any changes in the legal status of the drug [23]. For example, there is a clear relationship in the United States National Household Survey on Drug Abuse between perceived health risk and cannabis use from 1985 to 1997 among persons aged 12-17. One cannot be certain of the causal relationship without longitudinal data. Nonetheless, figure I shows that as perceived health risk increased in the late 1980s, rates of cannabis use declined. As perceived health risk diminished in the mid-1990s, rates of cannabis use began to rise again. This suggests that attitudes towards drug use play a key role in understanding trends in drug use. Attitudes are thus an important leverage point for drug prevention.

A further proximate cause of trends in drug use and its patterns is the extent to which there are alternatives to the illicit drug market. In countries where there are favourable cultural traditions and/or climatic conditions for the cultivation of illicit drug crops, changes in availability of drugs from the illicit drug market will have less impact on drug use as users can turn to home cultivation for their drug supplies. Alternative career and income prospects are a vital aspect in understanding why some persons become involved in illicit marketing. A major appeal of drug trafficking for some residents of urban ghettos in the United States is the fact that there are few opportunities in legitimate activities to achieve the level of income that dealing in illicit drugs can provide.

Underlying determinants: the role of drug policy

Environmental factors play a role in influencing the major proximate causes of drug use. Geographical isolation may inhibit or slow the spread of new drugs or new patterns of drug use. Isolation also simplifies the enforcement of laws prohibiting the import or export of illicit drugs. As noted earlier, climatic factors have an impact on the nature and extent of cultivation of illicit drug crops. Another critical factor concerning the environment is the extent to which AIDS and other

Figure I. Trends in cannabis use and perceived risk of cannabis use among persons aged 12-17 in the United States of America, 1985-1997



Source: E. Single, P. Christie and R. Ali. "The impact of cannabis decriminalization in Australia and the United States", *Journal of Public Health Policy*, vol. 21, 2000, pp. 157-186.

infectious diseases represent serious public health issues. In those countries where a substantial portion of the population is infected with communicable diseases that can spread via unsafe methods of injection drug use, there is a pressing need to institute risk reduction measures among injection drug users and thus there tends to be less support for purely abstinence-based approaches.

Nonetheless, the most critical underlying factor influencing the proximate causes of drug use and drug-related problems is drug policy, if only because this is the key leverage point in the model. A model developed for drug policy in British Columbia, Canada, is presented in figure II [24]. The framework components consist of the primary goal of the policy, strategies to achieve that goal, agencies responsible for interventions, strategic planning to develop programme priorities and performance indicators, research underpinning planning and evaluation, and funding. There is also a feedback loop in which performance indicators are monitored and the results are fed back into strategic planning for the next phase of the drug strategy.

The framework outlined in figure II applies to any drug policy, regardless of whether its main goal is to eliminate drug use or if its major objective is harm reduction. The strategies to achieve either goal fall into three general categories: demand reduction, supply interdiction and interventions directed at drug users. The three major strategies are not entirely mutually exclusive: improving treatment effectiveness reduces drug demand and reducing drug demand can affect aspects of drug supply. By the same token, impacts on supply can influence demand. For

example, Caulkins and colleagues have demonstrated how reductions in supply can increase the cost-effectiveness of drug law enforcement [25] and Moore has noted that drug law enforcement can enhance the effectiveness of treatment and prevention [26]. Nonetheless, these three strategies represent a reasonable classification of the major ways in which the goal of drug policy can be achieved.

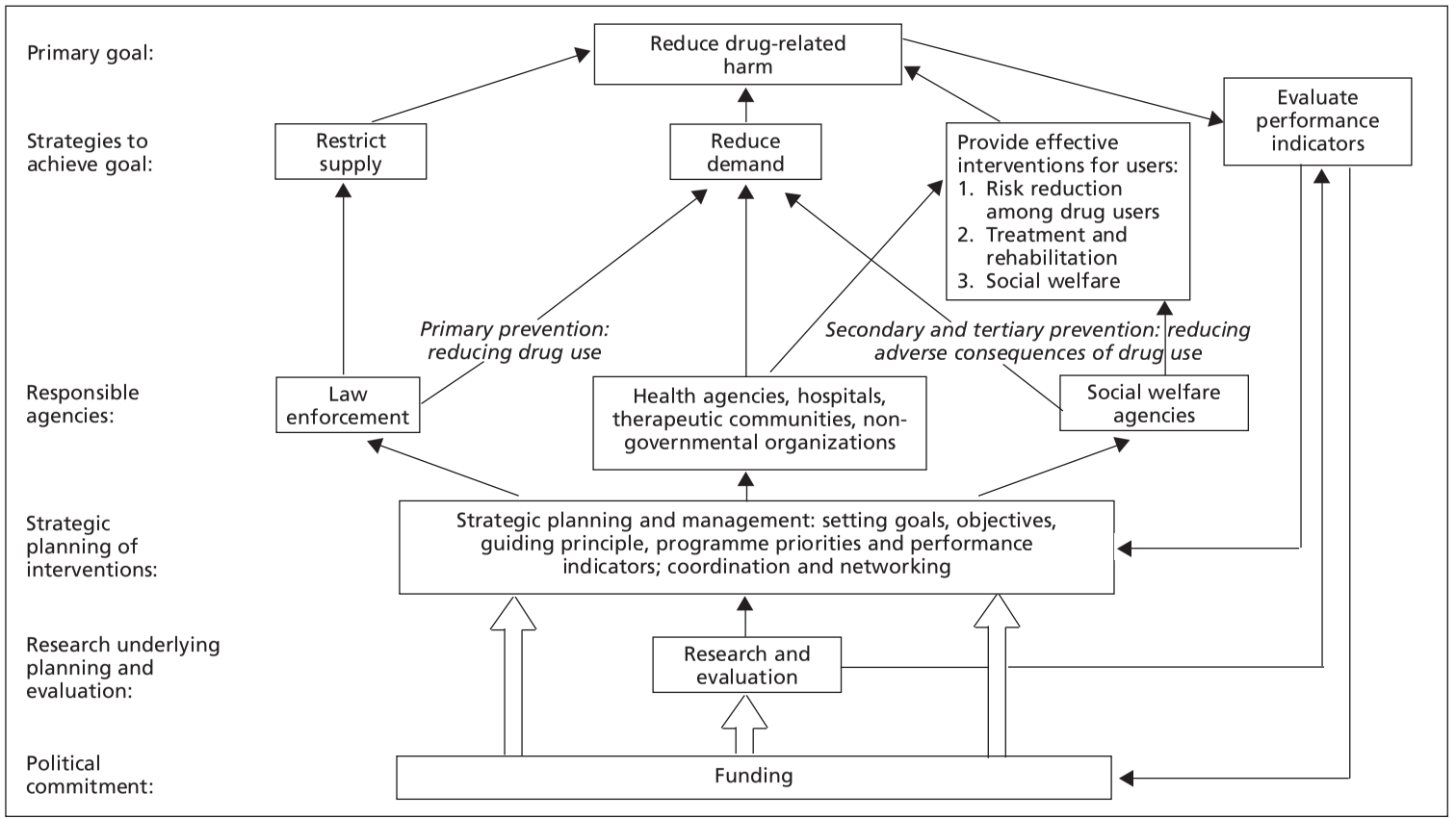
Interventions to achieve these objectives consist of prevention programming to reduce drug demand, the enforcement of drug laws and interventions directed at drug users. The first two sets of interventions—supply and demand reduction—correspond to primary prevention, while the third set—interventions directed at drug users—corresponds to secondary and tertiary prevention. Interventions aimed at drug users consist of risk reduction, treatment and rehabilitation of drug users, and social welfare policies that support treatment and rehabilitation. In a drug policy that focuses on reducing overall levels of drug use, little or no emphasis is placed on risk reduction measures such as syringe exchange and drug maintenance programmes. A drug policy based on harm reduction, on the other hand, places more emphasis on use-tolerant measures aimed at users who cannot be expected to cease their drug use at the present time.

Law enforcement is primarily responsible for supply interdiction and health agencies are generally responsible for demand reduction, while interventions aimed at drug users are the responsibility of both health and social welfare agencies. Again, there is necessarily some overlap with regard to responsible agencies. For example, treatment can reduce both illicit drug demand and/or supply, as when a treated user-seller stops using and dealing [26]. Similarly, law enforcement contributes to prevention programming by stigmatizing illicit drug use through school-based educational programmes and social welfare agencies contribute to the reduction of illicit drug demand.

Ideally, the three major types of interventions should be well planned and coordinated with one another. In practice, this is made difficult by the multiplicity of government and non-governmental organizations involved, for example, health officials from different levels of government, different police forces, hospitals, health-care workers, addiction treatment agencies and academic and non-academic researchers. To ensure effective strategic planning, goals should be agreed upon, as well as strategic objectives and guiding principles. Decisions must be made concerning programme priorities and funding, and performance indicators must be specified and monitored.

The basis of good strategic planning is research. In order to make sound decisions on programme priorities, it is necessary to have scientifically credible basic research on the basic biological mechanisms of dependence, the psycho-social risk factors and the interplay of individual characteristics, pharmacological properties of psychoactive substances and the environment in which consumption occurs. Applied research on the effectiveness of interventions is also vital. The requirements to evaluate a use-reduction drug policy are more limited than the requirements to evaluate a harm reduction policy—generally, trends in levels of drug use are all that is required to provide a broad evaluation of the former, while much more detailed information on specific drug-related harms is needed to assess a

Figure II. A conceptual framework for drug policy



Source: E. Single, "A harm reduction framework for drug policy in British Columbia", paper commissioned by the British Columbia Harm Reduction Working Group, Victoria, Canada, November 1999.

harm reduction approach. In either case, however, research plays a key role in the evaluation of performance indicators, which in turn provides information needed for strategic planning of interventions in the future.

Political commitment to the drug strategy is the final and perhaps the most essential component of the framework. The degree of political commitment determines funding levels, which dictates the limits of what can be accomplished. The framework is a dynamic model in that it includes a feedback loop whereby performance indicators are monitored and that information is used to adjust programming and strategically plan the next phase of the drug strategy. The results of evaluation also have an impact on the level of political commitment: a strategy that is achieving its goals is more likely to receive continued political support. A strategy that is unable to demonstrate its effectiveness is less likely to receive continued funding.

Summary and conclusions

The author of the present paper has presented a general model for understanding trends in drug use and drug-related problems over time in a particular setting. The major dependent variables of concern are rates of drug use, problematic patterns of drug use and indices of specific drug-related harm. The proximate causes of drug use and drug-related harm are drug availability, attitudes towards drugs and drug use, and the availability of alternatives to the use and marketing of illicit drugs. It is further posited that the major underlying factors that may influence the proximate causes include environmental factors that have an impact on drug crop cultivation and/or support for drug policies. Most importantly, drug policy is viewed as a major underlying determinant. Enforcement of drug laws, preventive education, treatment and other interventions with drug users represent the key leverage point by which trends in drug use and drug-related problems can be influenced by drug policy.

Unfortunately, the proposed model is subject to the criticism that it may be impractical and idealized, given the current state of knowledge. While development of a sound model for drug systems requires specification of the interrelationships of the major sets of variables in the model, an even more primary issue is the availability of data. There have been recent advances in improved measurement of problematic patterns of drug use (e.g. estimation of the number of injection drug users), drug-related harm (e.g. deaths and hospitalizations attributed to drugs, such as opioid dependence and overdose) and drug-related crime. Nonetheless, there is still a paucity of data on patterns of drug use or drug-related harm, so the viability of drug policy is generally measured in terms of changes in levels of drug use and/or changes in the number of persons detected and charged with drug crimes.

There is an oft-told anecdote about an inebriated man who searches for his dropped keys under a street lamp, not because the keys were lost there but because that is where the light is best. Like the drunk futilely looking for his keys

under the street lamp, undue focus has been placed on levels of use and drug charges as a surrogate for drug-related harm. Until valid and reliable data are available on all of its key variables, any model of drug systems will have limited applicability.

References

1. United Nations International Drug Control Programme, *World Drug Report* (London, Oxford University Press, 1997).
2. See, for example, P. Giffin, S. Endicott and S. Lambert, *Panic and Indifference: the Politics of Canada's Drug Laws* (Ottawa, Canadian Centre on Substance Abuse, 1991).
3. E. Single and T. Rohl, *The National Drug Strategy: Mapping the Future* (Canberra, Australian Government Publishing Service, 1997).
4. Department of Health, Housing, Local Government and Community Services, *National Drug Strategic Plan* (Canberra, Commonwealth Department of Health, Housing, Local Government and Community Services, 1993).
5. E. Single, *Performance Indicators for Canada's Approach to Substance Abuse* (Ottawa, Health Canada, 1997).
6. D. Collins and others, "Improving economic data to inform decisions in drug abuse control", *Bulletin on Narcotics,* forthcoming.
7. L. Robson and E. Single, *Literature Review of Studies on the Economic Costs of Substance Abuse* (Ottawa, Canadian Centre on Substance Abuse, 1995).
8. D. English and others, *The Quantification of Drug Caused Morbidity and Mortality in Australia, 1992* (Canberra, Commonwealth Department of Human Services and Health, 1995).
9. E. Single and others, "The economic costs of alcohol, tobacco and illicit drugs in Canada, 1992", *Addiction,* vol. 93, 1998, pp. 983-998.
10. See, for example, Correctional Services of Canada, *Final Report of the Working Group on Reducing Addictions* (Ottawa, Correctional Services of Canada, 1990).
11. S. Brochu and L. Guyon, "The issue of addiction among a sample of incarcerated women", 37th International Congress on Alcohol and Drug Dependence, 1995.
12. S. Brochu, "Estimating the costs of drug-related crime", Second International Symposium on Estimating the Social and Economic Costs of Substance Abuse, Montebello, Quebec, 1995.
13. W. Hall, J. Bell and J. Carless, "Crime and drug use among applicants for methadone maintenance", *Drug and Alcohol Dependence,* vol. 31, 1993, pp. 123-129.
14. B. Benson and others, "Is property crime caused by drug use or by drug enforcement policy?", *Applied Economics,* vol. 24, 1992, p. 690.
15. D. Kreuzer, "Drugs and delinquency", *European Criminology,* vols. 5 and 6, 1993.
16. R. Hammersley and others, "The relationship between crime and opioid use", *British Journal of Addiction,* vol. 84, 1989, pp. 1029-1043.
17. D. McBride and C. McCoy, "Crime and drug using behavior", *Criminology,* vol. 19, 1981, pp. 281-302.

18. J. Fagan, J. Weis and Y. Cheng, "Delinquency and substance use among inner city students", *Journal of Drug Issues,* vol. 20, 1990, pp. 351-402.

19. N. Brunelle et S. Brochu, "La prédiction de la délinquance et de la toxicomanie: les risques, les facteurs de risque", 63ième Congrès de l'ACFAS, 1995.

20. K. Pernanen, "The social cost of alcohol-related crime: conceptual, theoretical and causal attributions", Second International Symposium on Estimating the Social and Economic Costs of Substance Abuse, Montebello, Quebec, 1995.

21. J. Roth, *Psychoactive Substances and Violence* (Rockville, National Institute of Justice, United States Department of Justice, 1994).

22. A. Sutton and S. James, *Evaluation of Australian Drug Anti-trafficking Law Enforcement* (Canberra, National Policy Research Unit, 1996).

23. See, for example, E. Single, P. Christie and R. Ali, "The impact of cannabis decriminalization in Australia and the United States", *Journal of Public Health Policy,* vol. 21, 2000, pp. 157-186.

24. E. Single, "A harm reduction framework for drug policy in British Columbia", paper commissioned by the British Columbia Harm Reduction Working Group, Victoria, Canada, November 1999.

25. J. Caulkins and others, *Mandatory Minimum Drug Sentences: Throwing Away the Key or the Taxpayers' Money?* (Santa Monica, CA, RAND Corporation, 1997).

26. M. Moore, "Supply reduction and law enforcement", *Drugs and Crime,* M. Tonry and J. Wilson, eds. (Chicago, University of Chicago Press, 1990), pp. 109-158.

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

A mover-stayer type model for epidemics of problematic drug use

C. ROSSI*

Department of Mathematics, University of Rome "Tor Vergata", Italy

ABSTRACT

A modified version of a model for the human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) epidemic, proposed at the beginning of the 1990s and recently generalized, is presented as a representation of a problematic drug use epidemic. The model can be used both to estimate relevant epidemic macro-parameters and to carry out scenario analyses. The model is of the "mover-stayer" type and allows for heterogeneous risk behaviour among the susceptible population. Such models treat the "susceptibles" as subdivided into two groups: the "stayers", that is, the individuals who are considered not to be at risk of infection, and the "movers", who are at risk. Because of the interactions between infectious individuals (pushers) and the susceptibles, some of the latter may pass into the drug user "compartments" and begin a drug user "career". The model described here comprises two different stages of hidden drug use. The first (light use) stage, which can be more strictly defined, is the initial (or non-problematic) stage of drug use, following which light drug users can either stop using drugs or pass to hard drug use (or death). Other relevant stages taken into account are the therapy stage, the recidivist use stage (either light or hard use) and the temporary non-use stage (the latter are the visible stages). A simpler version of the model is studied using a Markov hybrid approximation and some "what if" scenario analyses are obtained by simulation. A more complex and realistic model is also outlined for possible further development.

Introduction: what only mathematical models can reveal

The drug problem and its consequences for society represent a complicated field for research. Policy makers and researchers are seeking answers to a number of questions concerning drug use, its consequences and related costs. They are examining the extent of drug use, the consequences of drug use, which policies are effective, the costs of such policies and the cost of drug use to society. The consequences comprise, among many others, the adverse effects of infectious diseases and the costs to society of drug-related criminality. It is thus important to under-

*The present article was funded in part by the European Monitoring Centre on Drugs and Drug Addiction in the framework of the European Network to Develop Policy Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions. The simulation procedure used to obtain scenario analyses was developed by Alessandro Ghizzoni.

stand and measure drug use and how it responds to drug control interventions [1]. The present article will contribute to that objective by introducing a simple compartmental epidemic model of drug use that incorporates various effects on initiation into new use or relapse into recidivist use of drugs.

Compartmental models represent a powerful mathematical tool that is well established in modelling the spread of diseases in a population. They provide a framework in which numbers of people in different "compartments" (each one homogeneous with respect to some specified characteristic) and the relationships between such compartments, modelling the dynamics of the population, can be described in mathematical terms.

Dynamic compartmental modelling of epidemic processes, producing either operational or transmission models, occurs through the usual representation of the dynamics involved by means of a system of stochastic or deterministic differential (or difference) equations; this is the case for both the operational and the transmission models. The main difference between the two kinds of model lies in the fact that transmission models take into account the dynamic processes at the microlevel, modelling the interactions between individuals belonging to the different subgroups involved in the epidemic, whereas operational models work on macro-variables or indicators suitable for use in estimating the size of the phenomena or monitoring the impact of various interventions, modelled by suitable scenario parameters. Many models of the two types have been developed to study the human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) epidemic and they can also be used, with some modifications, to represent problematic drug use epidemics.

Some examples of transmission models are given in Dietz [2], Hadeler [3], Kretzschmar and Dietz [4] in relation to the HIV/AIDS epidemic and in Billard and Dayananda [5] and in Behrens [1] for problematic drug use epidemics. Those models are complex because the researchers introduce into the modelling process a detailed formalization of the interactions existing, or assumed to exist, between a large number of subjects involved in the epidemic process. This is the case of the analysis of HIV transmission across risk groups, such as drug users and heterosexuals, when some specific hypothesis is formulated as regards the contact pattern [4, 6] or when the time at which the contact occurs is assumed somehow to affect the probability of transmission of the infection [4]. As already mentioned, such models can be helpful when specific contact or transmission patterns need to be thoroughly analysed, but they may be extremely cumbersome when the epidemic as a whole is being examined. Detailed model structures contain all sorts of parameters regulating every single interaction in the simulation process and seldom are enough data available for reliable estimation of all those many parameters.

A more efficient way of obtaining a picture of a simulated epidemic is by using a "simple" operational model. In contrast to transmission models, simple models do not attempt to include all the possible group or individual interactions into the modelling structure, but summarize the dynamics of the epidemic in terms of certain non-linear interactions and group the infected individuals in chains of "compartments". Most of the parameters controlling the dynamics in such systems are

derived from epidemiological studies, external to the model, and their values are based simply on specialized studies or on the literature; only a limited number of "internal" parameters are usually left to be estimated by fitting the existing data or to be used for scenario analyses. Typically, the set of internal parameters includes some form of control of the transmission and of the size of the core group; other internal parameters may have different origins and interpretations, depending on the design of each individual model. The ability of simple models to describe an epidemic correctly is theoretically much more limited when compared with that of complex transmission models. In general, however, simple models turn out to be much more tractable, both because of the limited number of parameters required for their functioning and because of the quality of their output. In fact, while the large number of parameters in the transmission models often results in their unidentifiability, parameter estimates in simple models are usually reliable enough for the models to be used successfully in complex scenario analyses and under other difficult conditions.

Most of the simple models have a stochastic basis, although their output results in some deterministic forecast curves. Such models, defined as "hybrid" models in Bailey [7-9], are the result of a deterministic approximation of stochastic models, using mean times of stay in compartments with conditional Poisson arrivals and maximum likelihood (ML) or numerical estimates of internal parameters. Their use, however theoretically disputable, is more and more frequent among applied researchers because of their technical simplicity and the clarity and immediate possibility of use and interpretation of their results: their output consists of deterministic curves of epidemic indicators (incidence, prevalence, dynamics of the core group and so on), while the parameter ML estimation procedure is derived directly from the distributional hypotheses of the original stochastic model. One such model for the HIV/AIDS epidemic was proposed at the beginning of the 1990s and has recently been generalized. It makes it possible to obtain scenario analyses with ease [10, 11]. That model, when used in conjunction with suitable back-calculation methods, can reduce uncertainties in the estimates of incidence curves [12] and has also been used for indirectly estimating the prevalence of injecting drug use in Italy [13]. A modified version of the model is presented below for use in relation to epidemics of problematic drug use. (The model will be presented from a mathematical point of view elsewhere [14].)

An operational model of problematic drug use epidemics

The graphs provided in figure I describe the main features of the proposed model. The model is of the "mover-stayer" type [10, 15, 16] and allows for heterogeneous risk behaviour among the susceptible population. Such models treat the susceptible population as subdivided into two groups: the "stayers", that is the individuals who are considered not to be at risk of infection (such models are suitable for scenario analyses in order to assess the impact of various proportions of vaccinated persons on the probability of extinction of a given epidemic) and the "movers", who are at risk. Because of the interactions between infectious

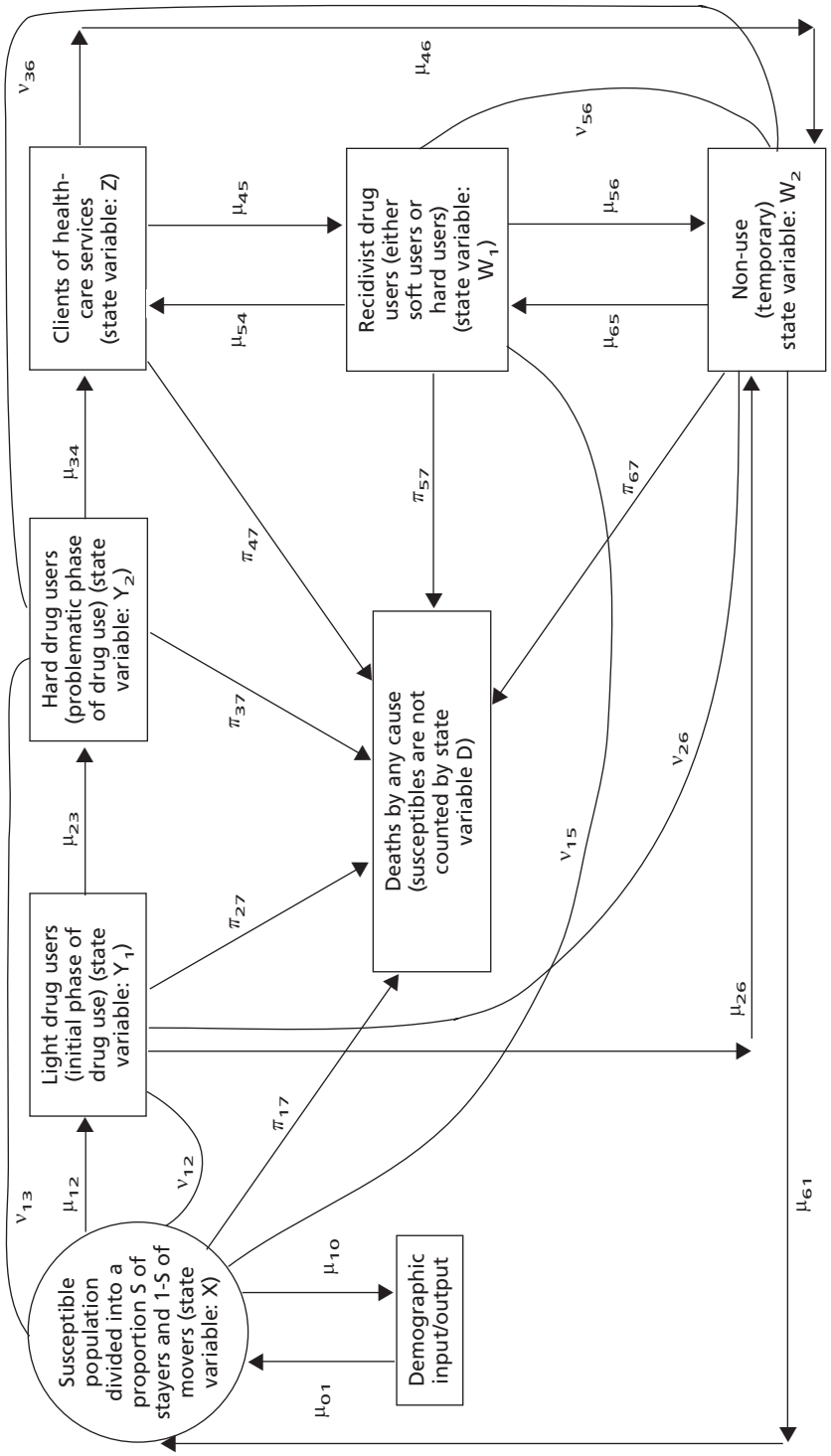
individuals (for the present study these are taken to be problematic drug users who are also pushers)* and the susceptibles, or as a result of the pressure of the black market on the susceptibles, some of the latter may pass into the drug user "compartments" and begin a drug user "career". As in the model proposed by Behrens and colleagues [1], the model described here comprises two different stages of hidden drug use. The first (light drug use) stage, which can be more strictly defined, is the initial (or non-problematic) stage of drug use, following which light drug users can either stop using drugs or pass to hard drug use (or death). The other arrows in the diagram show all the other possible transitions in a drug user career. The curves connecting the drug use (infectious) compartments and the susceptible (or temporary non-use) compartments indicate the possible interactions that may produce transitions from susceptibles (or temporary non-use) to infectious (non-linear terms in the equations); the other possible transitions from susceptibles (or temporary non-use) to infectious are induced by the pressure of the black market and are assumed to be represented in linear terms in the equations.

In order to formulate the corresponding equations (either deterministic or stochastic), some further hypotheses must be explored and the known and unknown parameters described. A first approximation may be through a Markov model, possibly a marked Markov process. In such a case the length of stay in each compartment is assumed to be distributed exponentially and the results are useful in providing a first qualitative insight into the epidemic process. A more realistic approximation is by using semi-Markov processes. In such a case the length of stay in each compartment may be assumed to be distributed differently with respect to an exponential variable, possibly depending also on covariates describing the individual experience of drug use. Such a study is much more complex, but suitable mathematical and simulation techniques can be used. In the present article, only the Markov model is considered and used to make some scenario analyses.

On the basis of figure I, it is immediately possible to formulate the equations of the model either in the form of deterministic (continuous or discrete) equations or as stochastic (continuous or discrete) equations. The discrete stochastic equations will be reported elsewhere [14]; only the simulation programme and some scenarios aimed at evaluating the impact of different parameters on the epidemic are considered in the present article. The state variables used in the model (with the exception of $S(t)$, which is the proportion of stayers at time t, represent the incidence per one million inhabitants; the unit time for the simulation runs is taken to equal one week. The equation for the proportion of stayers, S, is derived from the hypothesis [10] that the new entrances in the susceptible compartment are divided into stayers and movers according to constant proportions S_o and $1\text{-}S_o$ (stationarity), with $0 < S_o < 1$, even if other hypotheses can be incidentally incorporated into the model as well as possible transitions from the movers to the stayers compartments as a result of prevention campaigns or law enforcement activities. This further development is outlined briefly in the last part of this article.

*From surveys conducted among military conscripts, reported in the annual report on the state of the drug problem in Italy for the year 1999, published by the National Focal Point, it appears that, in terms of the reasons for drug use, the two most mentioned factors were curiosity (more than 40 per cent) and peer group pressure (more than 30 per cent).

Figure I. Compartmental representation of the system dynamic model of a problematic drug use epidemic



A qualitative analysis of the epidemic (transient and asymptotic behaviours), in particular as regards the incidence of drug use, that is the transitions from susceptible to drug user, can be conducted by analysing jointly the equation for X and the equation for S. The method is analogous to that used in Rossi [10] and will be considered elsewhere [14].

The crucial parameters to be estimated externally and the scenario parameters

As regards the parameters and the distributions of the length of stay, some are already available from study of the latency period [17]. Some can be derived using therapy data already available in various places. The demographic parameters regulating the dynamics of the susceptible population, μ_{01}, μ_{10} and π_{17}, are assumed to be known and are country-specific. The other μ parameters can be estimated externally using the information from mortality studies among drug users, which is available for most countries of the European Union [18]. The parameters μ_{23} and μ_{34} (natural history parameters) can be estimated on the basis of data available from study of the latency period [19, 20]. The parameters μ_{45}, μ_{46}, μ_{54}, μ_{56}, μ_{61} and μ_{65} (therapy parameters) can be obtained, at least as regards their order of magnitude, from therapy data available in most countries; the values of all these parameters for Italy (order of magnitude) are given in table 1. All the other parameters, μ_{12}, μ_{26}, ν_{12}, ν_{13}, ν_{15}, ν_{26}, ν_{36} and ν_{56}, can be used as scenario parameters, as well as the parameter "initial proportion of stayers", S_o.

In the present article it is assumed (following Billard and Dayananda [5]) that most pushers are hard drug users (basic ν-scenarios). It has been suggested elsewhere that most pushers are soft drug users [21], so some simulation runs will be devoted to obtaining scenarios based on this latter hypothesis (alternative ν-scenarios) by modifying the order of magnitude of the ν parameters reported in table 1 (10^{-5} instead of 10^{-6} and vice versa).

All the μ and π parameters represent transition rates per person of the origin compartment per week, the ν parameters are rates per week per pair. Using the simulation procedure described below, some impact analyses can be conducted to evaluate the influence of the scenario parameters on the course of the epidemic. It is also possible to make some further scenario analyses to evaluate the impact of increased efficacy of therapy services by taking the scenario parameters as fixed and using the therapy parameters as variable. Similarly, the natural history parameters can be used to make further "what if" scenario analyses (the results of such analyses will be presented elsewhere).

Some scenario analyses

The simulation procedure, used to obtain scenario analyses, is written in S-plus for personal computers. All the parameters can be modified at the beginning of each run. The standard output comprises graphs of the prevalence curves in each com-

Table 1. Transmission parameters estimated for Italy[a]

Between compartments	μ	π	ν *(order of magnitude)*
0-1	0.00025		
1-0	0.00002		
1-2	**$10^{-5}/10^{-6}$**		**10^{-6}**
1-3			**10^{-5}**
1-5			**10^{-6}**
6-1	0.0096		
1-7		0.00023	
2-3	0.009		
2-6	**0.004/0.0004**		**10^{-6}**
2-7		[0.0002-0.0008][b]	
3-4	0.004		
3-6			**10^{-5}**
3-7		[0.0002-0.0008][b]	
4-5	[0.014-0.018][c]		
4-6	[0.007-0.009][c]		
4-7		[0.0002-0.0008][b]	
5-4	0.001		
5-6	[0.05-0.1][d]		**10^{-6}**
6-5	0.001		
5-7		[0.0002-0.0008][b]	
6-7		[0.0002-0.0008][b]	

[a]Scenario parameters are shown in bold face (order of magnitude).

[b]Estimates from mortality studies used at the Consensus Conference on AIDS, held in Italy in 1998, reported by L. Ravà and others ("Estimating the size of the HIV/AIDS epidemic: complementary use of empirical Bayesian back calculation and the mover-stayer model for gathering the largest amount of information", *Simulation,* vol. 71, No. 4 (1998), pp. 213-227) and by the European Monitoring Centre for Drugs and Drug Addiction ("Coordination of implementation, follow-up and analysis of cohort studies on mortality among drug users in European Union Member States").

[c]Average length of treatment (hypothesis): 27-36 weeks.

[d]Average length of stay in compartment 5 (hypothesis): 5-10 weeks.

partment and of the incidence curves of major interest. It is possible to choose the total simulation time, which is measured in weeks.

Some scenario analyses have been conducted to study the impact of S_o and some other parameters on the qualitative behaviour of the epidemic. The results are reported in graphic form below. As expected, the effect of S_o is higher with respect to the other parameters [11] considered for the simulation runs reported here.

Basic ν-scenarios

Some scenarios have been obtained based on various hypotheses concerning the parameters. The results are reported in graphic form and a summary table (table 2) describes the macro-characteristics of the different epidemics using the location and the size of the peaks for the incidence and prevalence curves, the value of the cumulative curve of deaths at the end of the period taken into account and the location of the maximum of the proportion of stayers, which represents the time of the beginning of the endemic phase (saturation effect).

Scenario 1

Scenario 1 is obtained using the parameters listed in table 1, with S_o = 0.98 (this means that the size of the population at risk is taken to be about 20,000 per one million inhabitants), μ_{12} = 10^{-5} (this parameter measures the strength of the black market: its value means that the black market can be assumed to induce about one person of the risk group per month to try drugs at the beginning of the epidemic) and μ_{26} = 0.004 (this is the parameter regulating spontaneous exits from the non-addicted user population: its value means that the proportion of non-addicts who stop using drugs is about half those who proceed to addictive behavour). Various graphs representing the curves of interest are shown in figures II-VII.

Figure II. Scenario 1: Incidence curve from susceptibles to light users

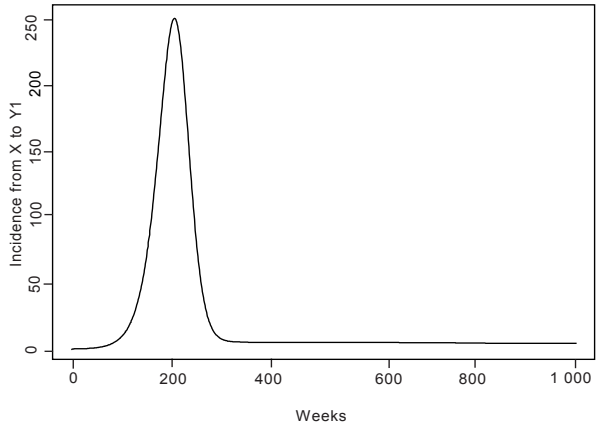


Figure III. Scenario 1: Prevalence curve of light users

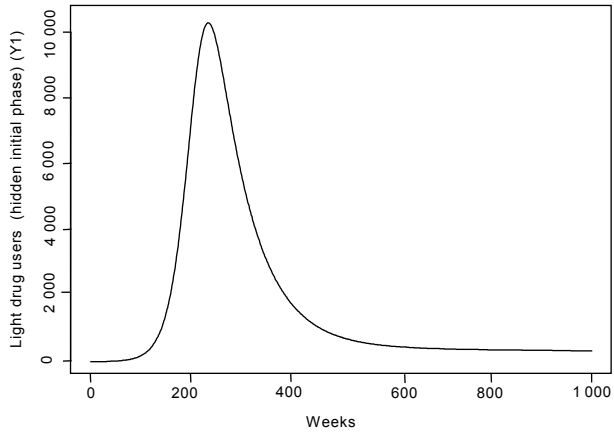


Figure IV. Scenario 1: Incidence curve from light users to heavy users

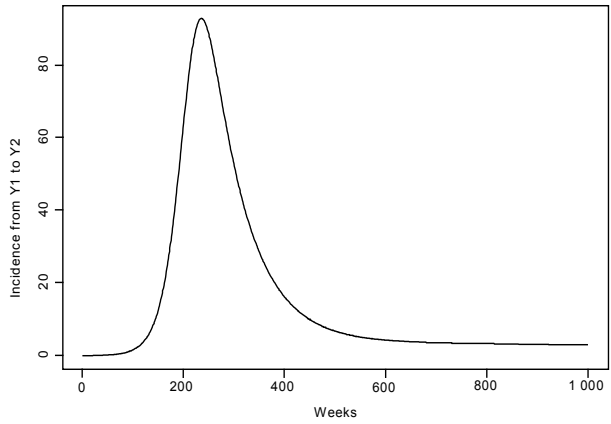


Figure V. Scenario 1: Prevalence curve of hard users

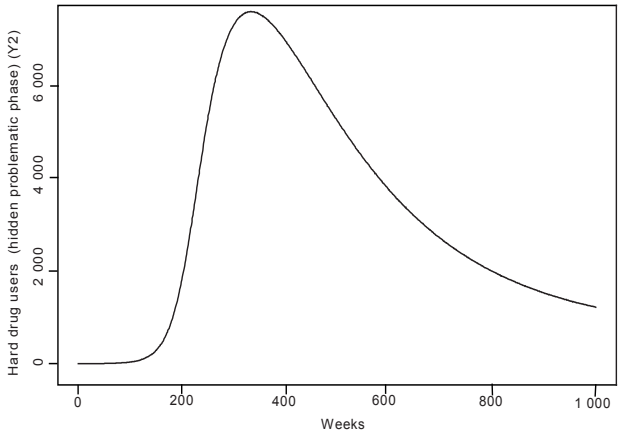


Figure VI. Scenario 1: Prevalence curve of clients of therapy services

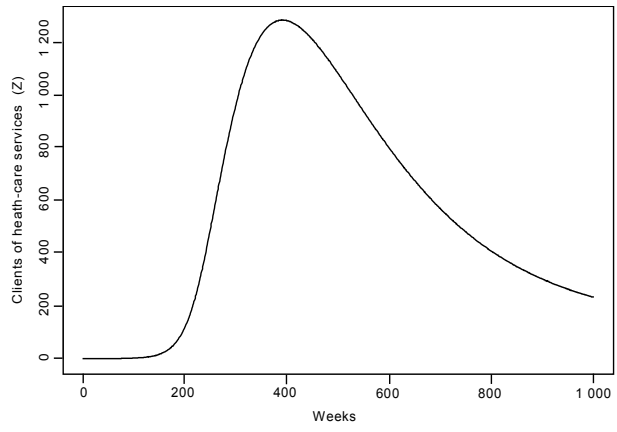
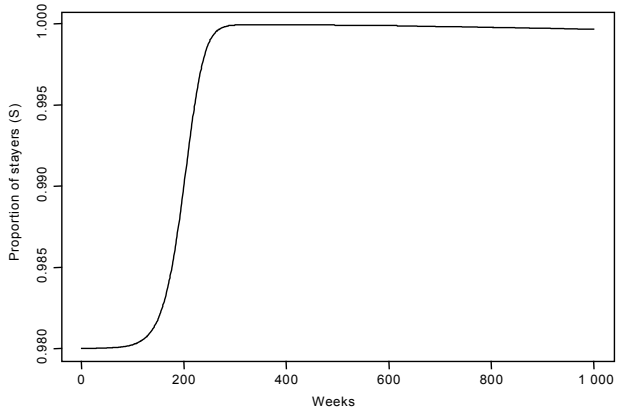


Figure VII. Scenario 1: Proportion of stayers



Scenario 2

This scenario is obtained using the parameters listed in table 1, with S_o = 0.99 (here the size of the population at risk is taken to be about 10,000 per one million inhabitants), μ_{12} = 10^{-5} (as in scenario 1) and μ_{26} = 0.004 (as in scenario 1). Graphs representing the curves of interest are shown in figures VIII-XIII.

Figure VIII. Scenario 2: Incidence curve from susceptibles to light users

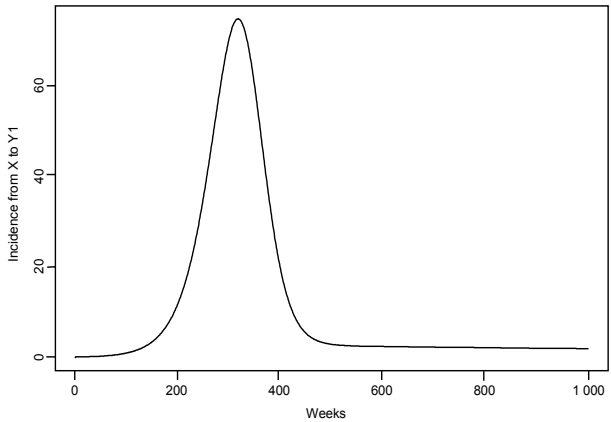


Figure IX. Scenario 2: Prevalence curve of light users

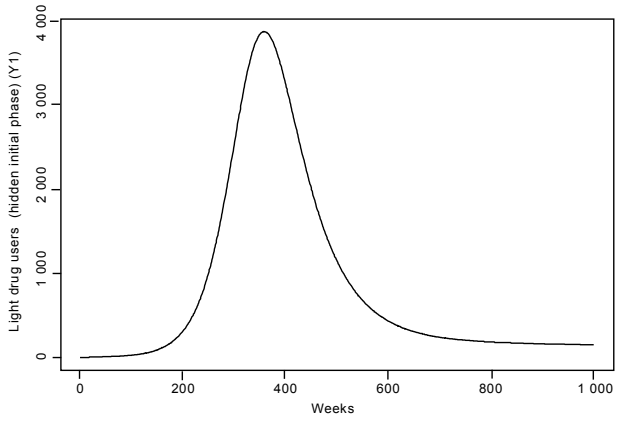


Figure X. Scenario 2: Incidence curve from light users to hard users

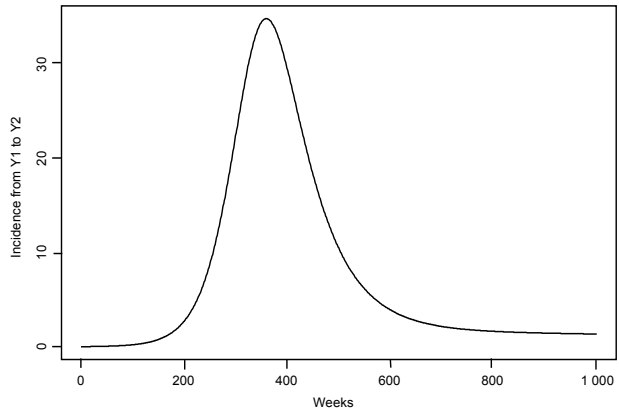


Figure XI. Scenario 2: Prevalence curve of hard users

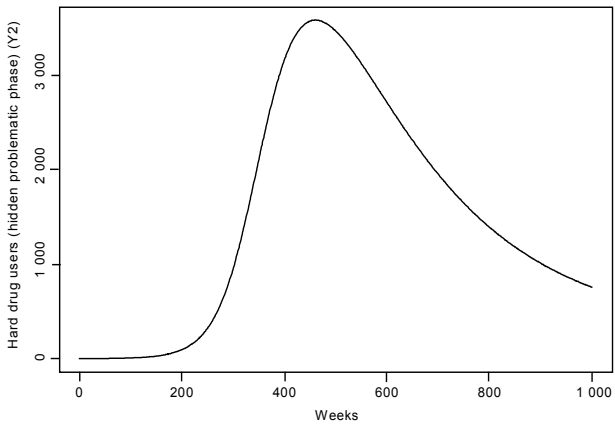


Figure XII. Scenario 2: Prevalence curve of clients of therapy services

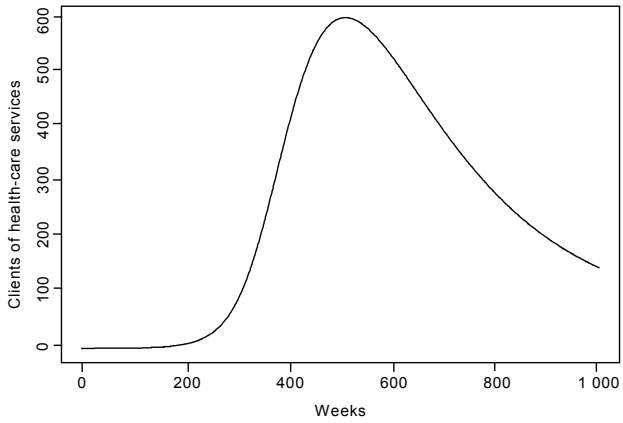
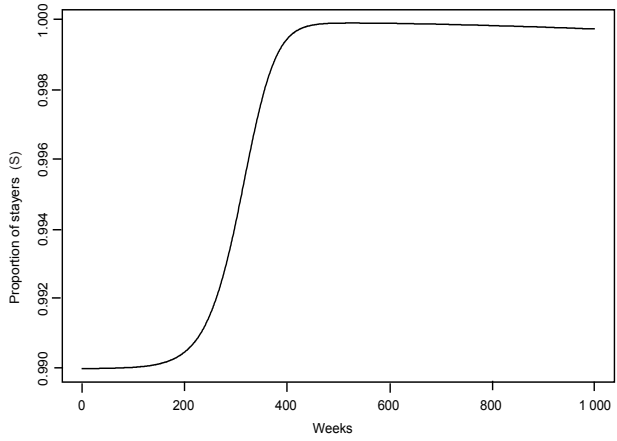


Figure XIII. Scenario 2: Proportion of stayers



Scenario 3

Scenario 3 is obtained using the parameters listed in table 1, with S_o = 0.98 (as in scenario 1), μ_{12} = 10^{-6} (this parameter measures the strength of the black market: its value means that the black market can be assumed to induce about one person of the risk group per 10 months to try drugs at the beginning of the epidemic) and μ_{26} = 0.004 (as in scenario 2). Graphs representing the curves of interest are shown in figures XIV-XIX.

Figure XIV. Scenario 3: Incidence curve from users

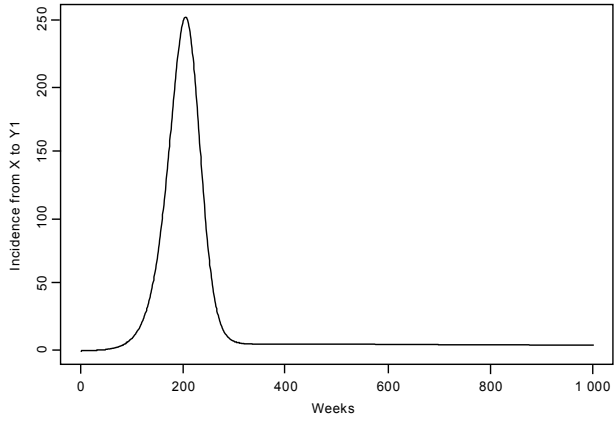


Figure XV. Scenario 3: Prevalence curve of light susceptibles to light users

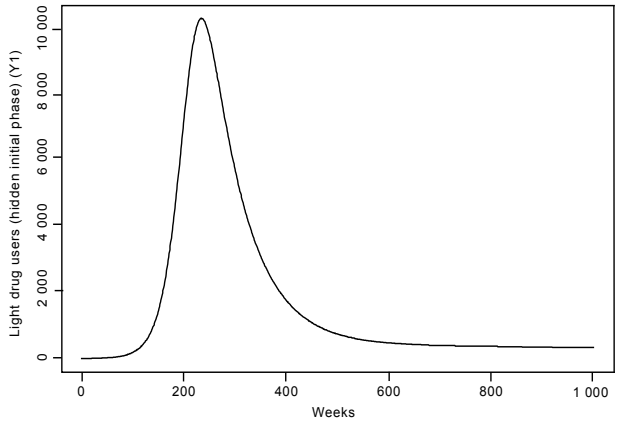


Figure XVI. Scenario 3: Incidence curve from light users

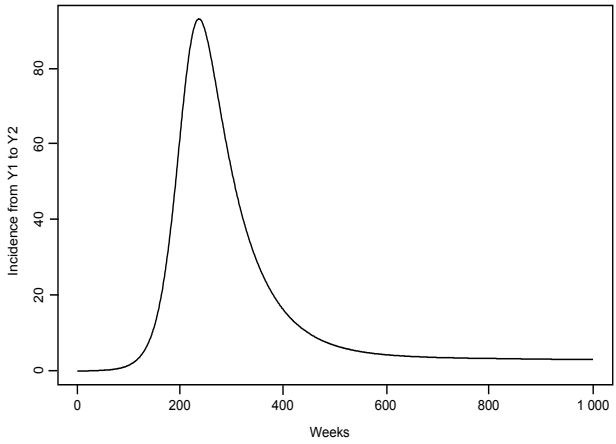


Figure XVII. Scenario 3: Prevalence curve of hard users to light users

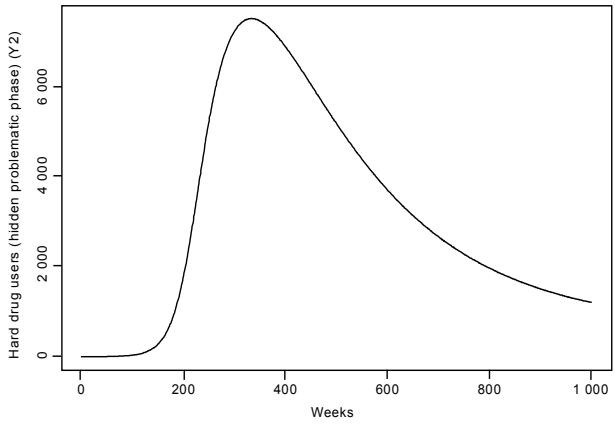


Figure XVIII. Scenario 3: Prevalence curve of clients of therapy services

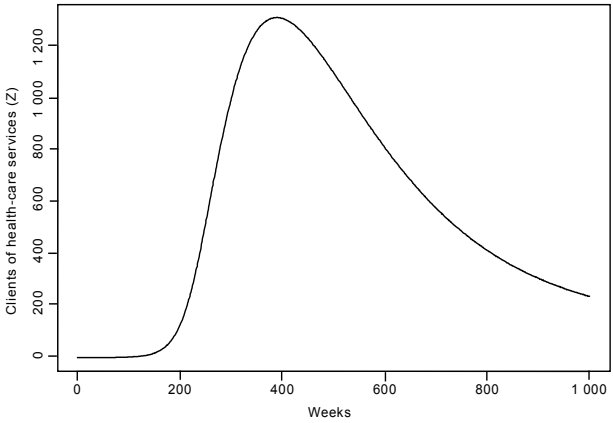
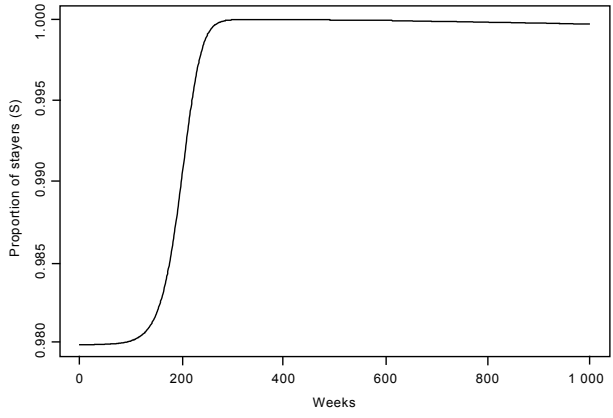


Figure XIX. Scenario 3: Proportion of stayers



Scenario 4

This scenario is obtained using the parameters listed in table 1, with S_o = 0.98 (as in scenario 3), μ_{12} = 10^{-5} (as in scenarios 1 and 2) and μ_{26} = 0.0004 (this is the parameter regulating spontaneous exits from the non-addicted user population: its value means that the proportion of non-addicts who stop using drugs is about 1/20 of those who proceed to addictive behaviour). Graphs representing the curves of interest are shown in figures XX-XXV.

Figure XX. Scenario 4: Incidence curve from susceptibles to light users

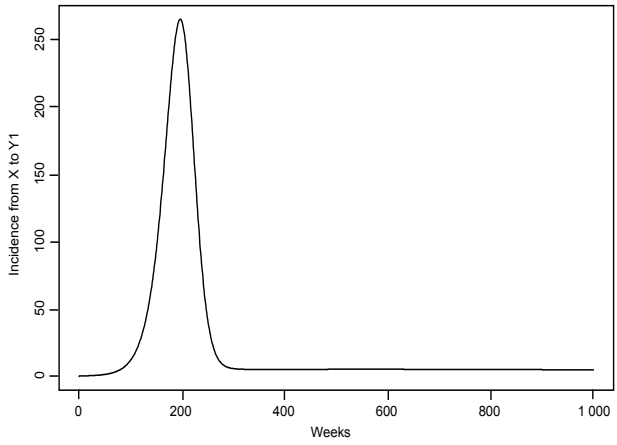


Figure XXI. Scenario 4: Prevalence curve of light users

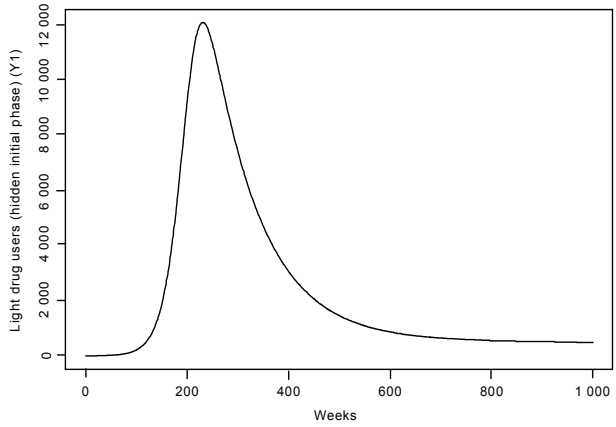


Figure XXII. Scenario 4: Incidence curve from light users to hard users

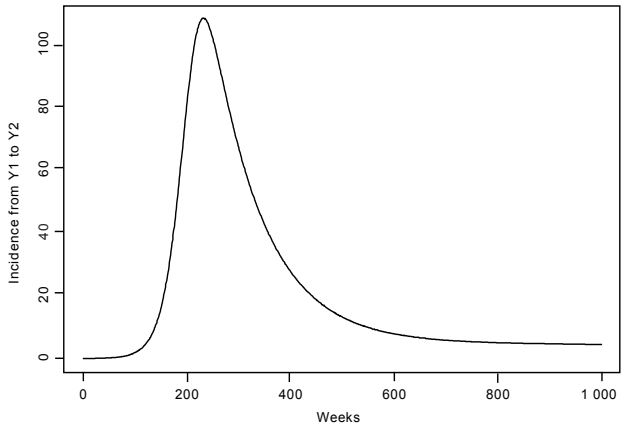


Figure XXIII. Scenario 4: Prevalence curve of hard users

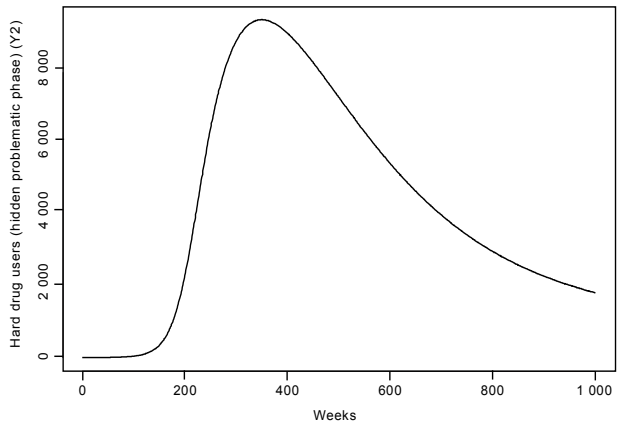


Figure XXIV. Scenario 4: Prevalence curve of clients of therapy services

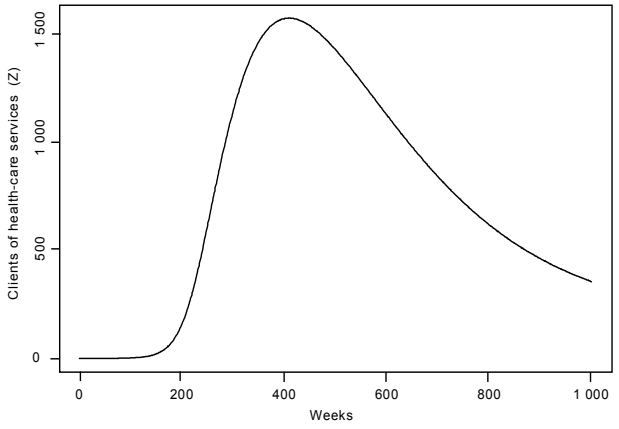
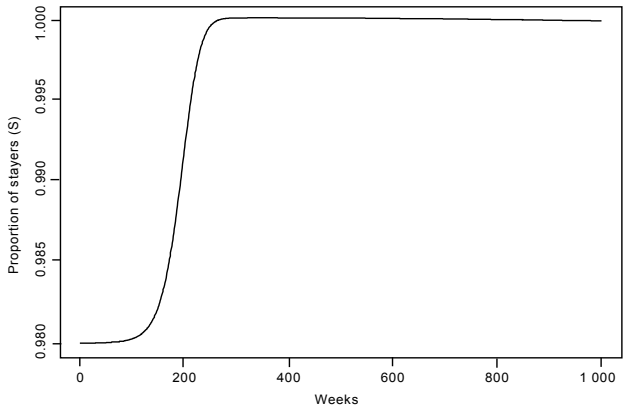


Figure XXV. Scenario 4: Proportion of stayers



Scenario 5

The present scenario is obtained using the parameters listed in table 1, but modifying the order of magnitude of the ν parameters (10^{-5} instead of 10^{-6} and vice versa). This means that it is assumed that there is higher possibility that a non-addicted individual can induce a susceptible to try drugs than an addicted one can. The values S_o = 0.98, μ_{12} = 10^{-5} and μ_{26} = 0.004 are the same as in scenario 1. Graphs representing the curves of interest are shown in figures XXVI-XXXI.

Figure XXVI. Scenario 5: Incidence curve from susceptibles to light users

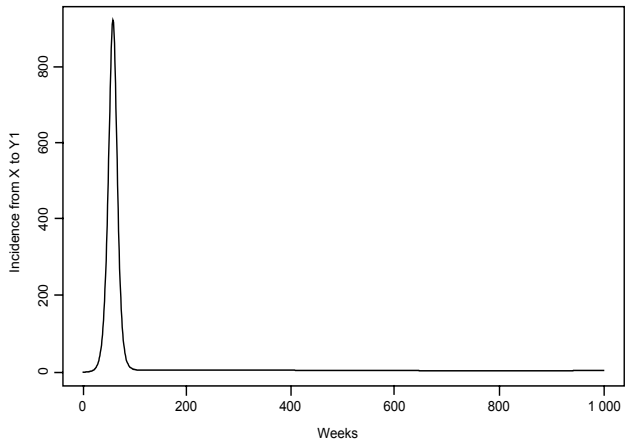


Figure XXVII. Scenario 5: Prevalence curve of light users

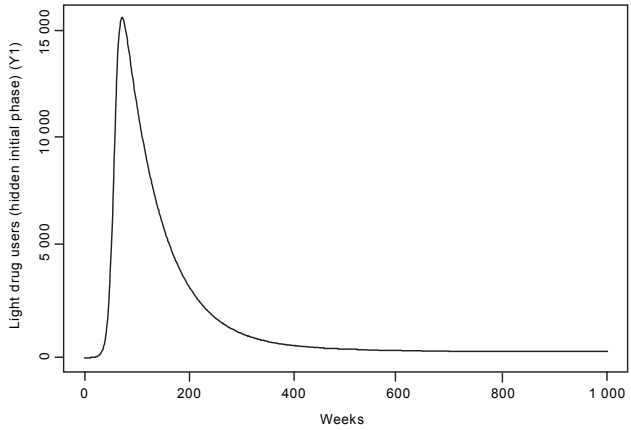


Figure XXVIII. Scenario 5: Incidence curve from light users to hard users

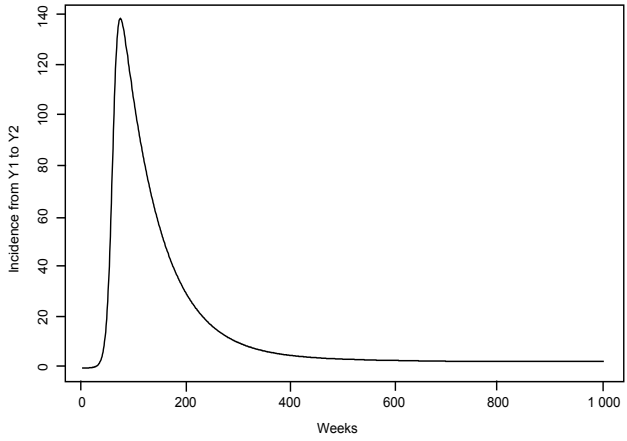


Figure XXIX. Scenario 5: Prevalence curve of hard users

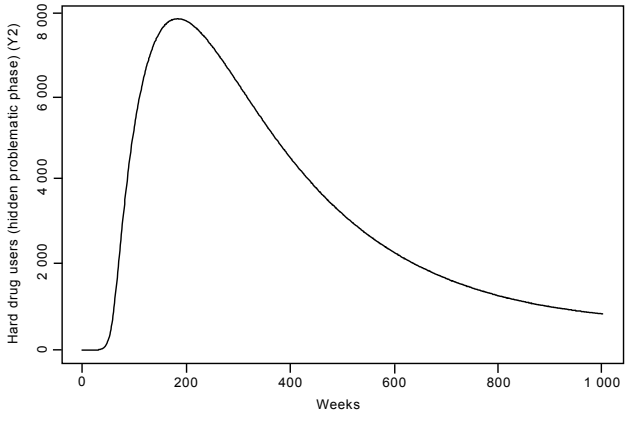


Figure XXX. Scenario 5: Prevalence curve of clients of therapy services

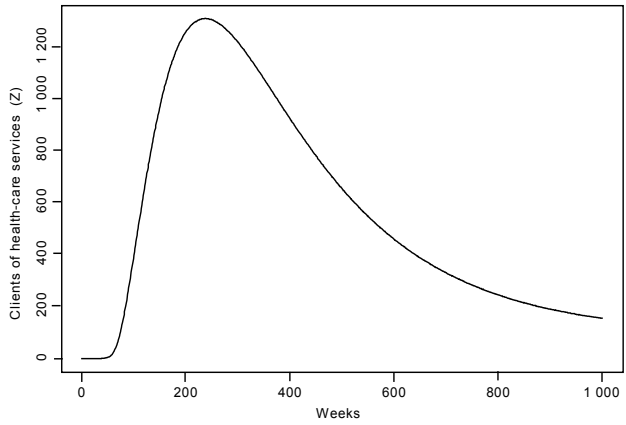
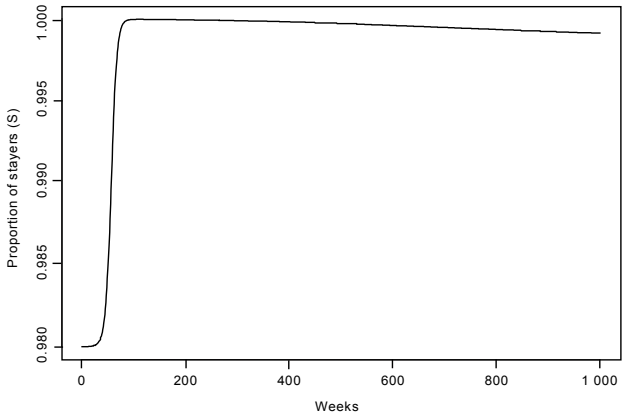


Figure XXXI. Scenario 5: Proportion of stayers



Scenario 6

The present scenario is obtained using the parameters listed in table 1, but modifying the order of magnitude of the ν parameters (10^{-5} instead of 10^{-6} and vice versa), as in scenario 5, and with S_o = 0.99, μ_{12} = 10^{-5} and μ_{26} = 0.004 (as in scenario 2). Graphs representing the curves of interest are shown in figures XXXII-XXXVII.

Figure XXXII. Scenario 6: Incidence curve from susceptibles to light users

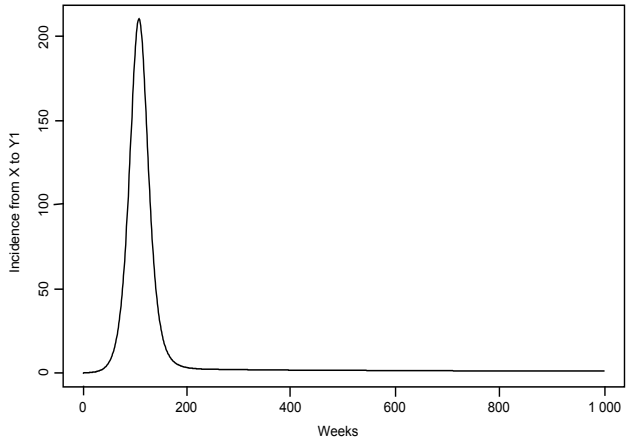


Figure XXXIII. Scenario 6: Prevalence curve of light users

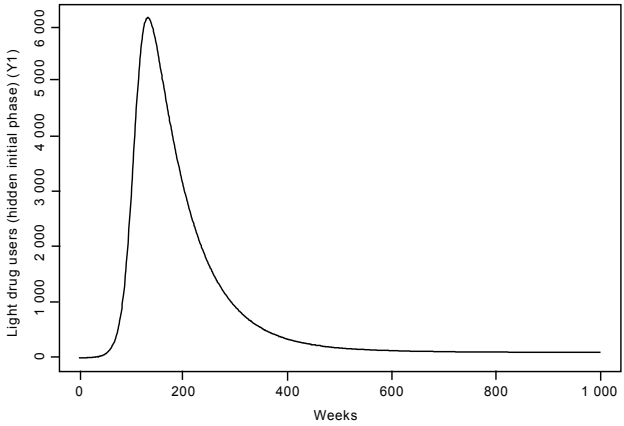


Figure XXXIV. Scenario 6: Incidence curve from light users to hard users

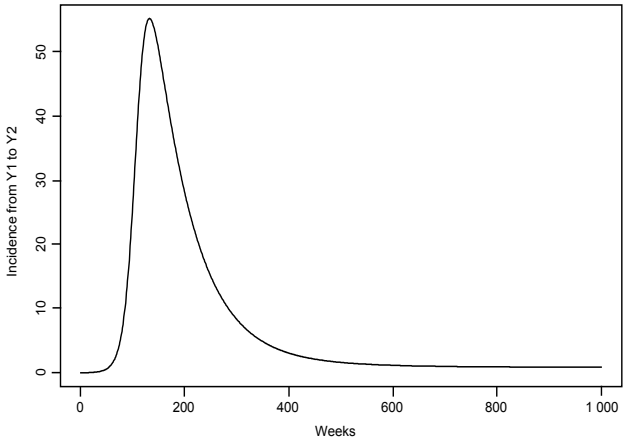


Figure XXXV. Scenario 6: Prevalence curve of hard users

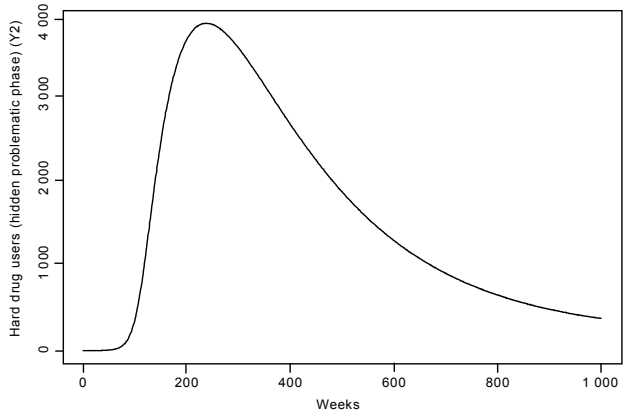


Figure XXXVI. Scenario 6: Prevalence curve of clients of therapy services

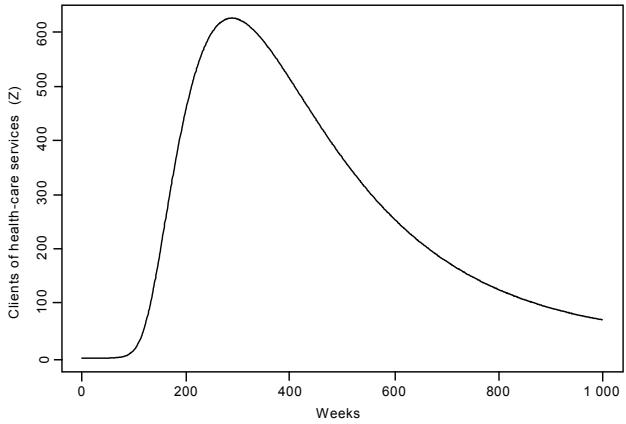
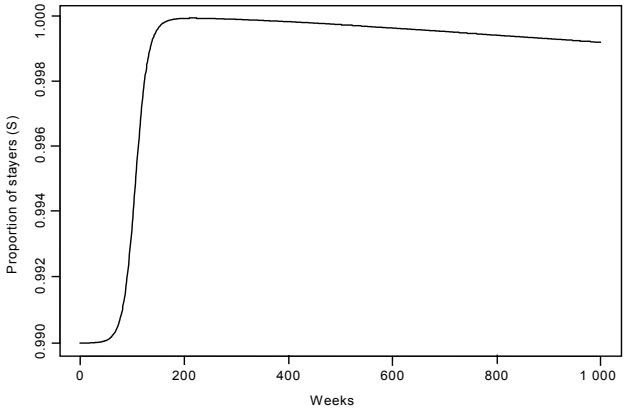


Figure XXXVII. Scenario 6: Proportion of stayers



An analysis of the graphs shows clearly that the parameter with the highest impact on the course of the epidemic is S_o, which is a measure of the size of the group of susceptibles who are at risk of infection (the core group). The bigger the core group, the faster the evolution of the epidemic and the higher the prevalence and incidence curves. The influence of the parameter that measures the pressure of the black market, μ_{12}, appears to be less important. As regards the comparison between the basic hypothesis for the ν parameters, that is, that pushers are mainly hard users, and the alternative hypothesis, that pushers are mainly light users, on the basis of the results of the simulations, it appears that the alternative scenarios show faster evolutions of epidemics. Table 2 summarizes the results in terms of the macro-indicators (location and size of the peaks for the incidence and prevalence curves, value of the cumulative curve of deaths at the end of the period taken into account and the location of the maximum of the proportion of stayers) described above. All the values related to size represent rates per one million inhabitants.

It can be observed that the initial proportion of stayers, S_o is the parameter with the highest impact: there are higher differences in the macro-parameters between scenario 1 and scenario 2 than between scenarios 1 and 3 or between scenarios 1 and 4.

Higher values of S_o correspond to larger and faster-growing epidemics, whereas the parameter related to the pressure of the market is not as important, at least in the range of values examined here. A minor effect also arises from parameter μ_{26}, which models transitions from light use to non-use (spontaneous cessation). As mentioned above, alternative scenarios correspond essentially to faster-growing epidemics, other things beings equal.

Further analyses will be reported elsewhere [14].

Further developments

The use of suitable markers (marked processes) might make it possible to incorporate further descriptions of each individual involved in an epidemic; for instance, markers may take into account the number of incarcerations, the number of non-fatal overdoses or the number of failed therapy interventions. Possible repeated therapy interventions should be incorporated into the model in order to obtain a more realistic picture of the career of the problematic drug user; there is general agreement that the time spent in the therapy compartment for the first therapy episode is distributed differently with respect to times related to the following episodes. The complexity of such analysis is evident and the interpretation of the results may be quite uncertain and unreliable owing to lack of information about the statistical distribution of the length of stay in the various compartments. It is therefore preferable to wait until the therapy data sets comprise reliable and complete information that will make possible external estimation of the parameters of interest.

Table 2. Macro-indicators describing the results of the various scenario analyses
(L = location of the peak; S = size of the peak)

	Basic scenarios								*Alternative scenarios*			
	1		*2*		*3*		*4*		*5*		*6*	
	L	*S*	*L*	*S*	*L*	*S*	*L*	*S*	*L*	*S*	*L*	*S*
Prevalence and incidence curves[a]												
Incidence curve from susceptibles to light users	50	250	75	80	50	250	50	250	15	900	30	210
Prevalence curve of light users	62	10 000	90	3 800	62	10 000	62	12 000	20	1 500	40	6 200
Incidence curve from light users to hard users	62	90	90	35	62	90	62	120	20	140	40	60
Prevalence curve of hard users	80	7 500	125	3 500	80	7 500	80	9 000	50	8 000	60	3 800
Incidence curve from hard users to therapy	90	30	110	15	90	30	90	38	50	30	60	16
Prevalence curve of clients of therapy services	100	1 200	125	600	100	1 200	100	1 600	62	1 300	64	620
Prevalence curve of recidivist use	100	2 700	125	800	100	2 700	100	3 000	62	2 500	64	700
Prevalence curve of no use (temporary)	130	3 000	150	1 500	130	3 000	150	2 600	89	3 200	80	1 700
Cumulative death curve		4 100		1 800		4 100		5 000		4 100		1 800
Proportion of stayers[b]	70		100		70		62		18		40	

[a]The unit time for the incidence curves is one week; the prevalence curves represent the sizes concerning cross-sectional observations.

[b]The proportion of stayers is an increasing function until the epidemic reaches the endemic phase and a decreasing function afterwards (C. Rossi, "A stochastic mover-stayer model for HIV epidemics", *Mathematical Biosciences*, vol. 107, 1991, pp. 521-545; and C. Rosse and G. Schinaia, "The mover-stayer model for the HIV/AIDS epidemic in action", *Interfaces*, vol. 28, No. 3 (1998), pp. 127-143).

In order to obtain more realistic models, the transition parameters should not be taken as constant, but should be represented as functions taking into account the history of drug use for a given individual, represented by statistical variables taken as known (covariates), and the history of policy interventions (availability of services, law enforcement activities and so on), represented by other covariates (time-dependent) and, possibly, by latent variables. This would result in a realistic but at the same time unreliable and intractable transmission model of very limited use. Some years ago, Kaplan [22] published a paper entitled "Can bad models suggest good policies?" in which he clearly explained how simple models, taking into account only the main peculiarities of the phenomenon of interest and neglecting minor effects, which may simply mask the relevant behaviours, are much more useful in order to assess the main consequences of policy interventions than complex models. The title itself is a good indication of the usefulness of simple mathematical models in addressing complex policy issues. Finally, if it were to be necessary to evaluate the impact of some kind of prevention campaign or intervention relating to problematic drug use among susceptibles, using the mover-stayer model proposed above, that issue could be dealt with easily by introducing the possibility that a mover become a stayer during the period of the campaign (or soon after). This can be simply modelled by assuming that during a short period a given proportion of movers become stayers and modifying the equations of the model to incorporate that fact. Analysis of the new equations would then make it possible to state that the efficacy of a primary prevention intervention of that length is higher at the beginning of the epidemic and decreases in the following period [14]. The effect of a secondary prevention intervention can also be taken into account in a similar manner by modifying the equations of the model. Analysis of the new equations would make it possible to state that the efficacy of a secondary prevention intervention of the given length is lower at the beginning of the epidemic and increases in the following period. The model may thus always be used to forecast the effect of such types of intervention and to determine the best combination of the two [14]. Similarly, the effect of law enforcement, which may influence movers, pushing a proportion of them to become stayers as a result of information about the adverse consequences of using illegal drugs [1], can be modelled assuming some parameters depending either on the number of addicts assisted by the health-care services or the number of incarcerated addicts or both.

Further generalizations might concern a more realistic approach to modelling the length of stay in the various compartments, taking into account the heterogeneity of individual behaviours. Most of these issues will be addressed in future contributions.

References

1. D. A. Behrens and others, "A dynamic model of drug initiation: implications for treatment and drug control", *Mathematical Biosciences,* vol. 159, No. 1 (1999), pp. 1-20.
2. K. Dietz, "On the transmission dynamics of HIV", *Mathematical Biosciences,* vol. 90, 1988, pp. 397-414.

3. K. P. Hadeler, "Modeling AIDS in structured populations", in *Proceedings of I.S.I. 47th Session,* Paris, 1989, pp. 83-99.
4. M. Kretzschmar and K. Dietz, "The effect of pair formation and variable infectivity on the spread of an infection without recovery", *Mathematical Biosciences*, vol. 148, No. 1 (1998), pp. 83-113.
5. L. Billard and P.W.A. Dayananda, "Drug addiction—pushers generated from addicts", *Biometrical Journal,* vol. 35, 1993, pp. 227-244.
6. J. A. Jacquez and others, "Modeling and analyzing HIV transmission: the effect of contact patterns", *Mathematical Biosciences,* vol. 92, 1988, pp. 119-199.
7. N. T. J. Bailey, "The use of operational modelling of HIV/AIDS in a system approach to public health decision making", *Mathematical Biosciences,* vol. 107, 1991, pp. 413-430.
8. N. T. J. Bailey, "An improved hybrid HIV/AIDS model geared to specific public health data and decision making", *Mathematical Biosciences,* vol. 117, 1993, pp. 221-237.
9. N. T. J. Bailey, "HIV/AIDS core-group dynamic and public health action", in *Modeling the AIDS epidemic,* E. H. Kaplan and M. L. Brandeau, eds. (New York, Raven Press, 1994), pp. 29-52.
10. C. Rossi, "A stochastic mover-stayer model for HIV epidemics", *Mathematical Biosciences*, vol. 107, 1991, pp. 521-545.
11. C. Rossi and G. Schinaia, "The mover-stayer model for the HIV/AIDS epidemic in action", *Interfaces,* vol. 28, No. 3 (1998), pp. 127-143.
12. L. Ravà and others, "Estimating the size of the HIV/AIDS epidemic: complementary use of empirical Bayesian back calculation and the mover-stayer model for gathering the largest amount of information", *Simulation,* vol. 71, No. 4 (1998), pp. 213-227.
13. C. Rossi, "Estimating the prevalence of injecting drug users on the basis of Markov models of the HIV/AIDS epidemic: applications to Italian data", *Health Care Management Science,* vol. 2, 1999, pp. 173-179.
14. C. Rossi, "Operational models for problematic drug use epidemics: the mover-stayer approach to heterogeneity", forthcoming.
15. T. M. Morgan, C. S. Aneshensel and V. A. Clark, "Parameter estimation for mover-stayer models: analysing depression over time", *Sociological Methods and Research,* vol. 11, 1983, pp. 345-366.
16. P. S. Albert, "A mover-stayer model for longitudinal marker data", *Biometrics,* vol. 55, No. 4 (1999), pp. 1252-1257.
17. European Monitoring Centre for Drugs and Drug Addiction, "Pilot project to estimate time trends and incidence of problem drug use in the European Union", final report, Lisbon, 1999.
18. European Monitoring Centre for Drugs and Drug Addiction, "Coordination of implementation, follow-up and analysis of cohort studies on mortality among drug users in European Union Member States", final report, Lisbon, 1999.
19. C. Rossi, "Monitoring drug control strategies: hidden phenomena, observable events, observable times", International Journal of Drug Policy, vol. 10, No. 1 (1999), pp. 131-144.

20. L. Ravà and others, "The use of back-calculation for investigating temporal-gegraphic spread of problem drug use: nationwide and regional analysis in Italy", forthcoming.

21. L. G. Hunt and C. D. Chambers, *The Heroin Epidemics* (New York, Spectrum Publications, 1976).

22. E. Kaplan, "Can bad models suggest good policies? Sexual mixing and the AIDS epidemic", *Journal of Sex Research,* vol. 26, 1989, pp. 301-314.

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

The dynamic process of dynamic modelling: the cocaine epidemic in the United States of America*

D. A. BEHRENS
Department of Economics, University of Klagenfurt, Klagenfurt, Austria

G. TRAGLER
Department of Operations Research and Systems Theory, Vienna University of Technology, Vienna, Austria

ABSTRACT

In the present article, the authors review several recent dynamic models of the current cocaine epidemic in the United States of America (both uncontrolled and optimally controlled), which differentiate between two levels of use ("light" and "heavy"). Even though all the models have their origin in a study carried out at the RAND Corporation's Drug Policy Research Center in the early 1990s, each has been developed by extending or refining another. In addition to pointing to interesting policy conclusions drawn from the analysis of those models, the authors also demonstrate that the development of dynamic models of illicit drug consumption is itself a dynamic process where subsequent refinements lead to increased quality and reliability of the resulting policy conclusions.

Introduction

Illicit drug use and related crime have imposed significant costs on the United States of America and various source and trans-shipment countries for a number of years. A variety of control strategies exist, including prevention, treatment and various forms of enforcement, so a fundamental question in drug policy is how scarce resources should be allocated between the various programmes. Analysts have sought to inform that decision by estimating the cost-effectiveness of different interventions. The greater part of that work has made estimates only for a particular point in time, concluding, for example, that in 1992 domestic enforcement was three times more cost-effective than border interdiction [1]. Earlier studies that used dynamic models have not focused on cost-effectiveness [2-6].

Behrens and others [7] contributed to the effort of understanding drug use and how it responds to drug control interventions by introducing a simple continuous time model of drug demand that incorporates a feedback effect of the

*The present research was financed in part by the Austrian Science Foundation (FWF).

current prevalence, or level, of use on initiation into new use. Analysing the model generates important new insights into how epidemics of drug use should be studied and, to the extent that such a simple model can be trusted, how they should be controlled. The model has been subject to refinements along different lines.

Firstly, in the original model, the deterrent effect on initiation was governed by the current number of heavy users. That approach kept the model's complexity low enough to simplify the analysis, while featuring the rather unrealistic assumption that when a heavy user exited the user population, his or her contribution to deterring initiation went immediately to zero. In other words, there was absolutely no memory of past bad experiences with heavy use even though in reality not all knowledge of a heavy user's bad experiences disappears the moment the individual exits the population, in particular if the exit is by death from drug use. More recent models avoid that problem by introducing a third category (in addition to the numbers of light and heavy users) that reflects some sort of memory of drug abuse.

In the present paper the model of Behrens and others [7] is reviewed, as well as the family of models stemming from its refinements [8, 9]. In addition to a description of those dynamic models, major policy conclusions are drawn on the basis of the analyses made. Hence, the purpose here is twofold. In an application-oriented vein, concrete examples are provided of dynamic models that have been parameterized with data concerning the current cocaine epidemic in the United States. In a more philosophical vein, reference is made to the fact that the development of dynamic models is itself a highly dynamic process owing to the fact that data sources are continuously reaching higher quality, understanding of the underlying epidemic processes is growing and the recent development of the tools needed for the analysis of dynamic models (both hard- and software) has led to subsequent improvements of the models (and in turn to higher quality/reliability of the resulting policy conclusions).

A dynamic model of the cocaine epidemic in the United States, including endogenous feedback on initiation

Since there is enormous heterogeneity between drug users with respect to rates of consumption and since the average rate of consumption for a population can change over time, tracking trends in total consumption (which is a reliable measure for the size of a drug problem, according to Rydell and others [1]) requires separate modelling of the numbers of users at different levels, or intensities, of drug use. Ideally, the whole spectrum of consumption behaviour would be modelled, from occasional use in small amounts up to frequent use in large amounts, but data limitations make that infeasible.

Everingham and Rydell [10] recognized that tension and suggested that, at least for cocaine, a simple dichotomous distinction between "light" and "heavy" users was useful. They operationalized the distinction using data from the *National Household Survey of Drug Abuse* [11], which measure the prevalence of cocaine use among the household population in the United States. In particular, people

who reported using cocaine "at least weekly" were defined as "heavy" users, while those who had consumed at least once within the last year but had done so less than weekly were called "light" users. The average heavy user consumes cocaine at a rate approximately seven times that of an average light user and exhibits substantially greater adverse consequences associated with that drug use.

A significant limitation of Everingham and Rydell's model [10] was that initiation was scripted. Future projections and policy simulation exercises were predicated on a fixed projection of future initiation that was insensitive to the course of the drug epidemic. This is problematic because the current prevalence of use significantly influences initiation rates. In particular, most people who start using drugs do so through contact with a friend or sibling who is already using drugs. Indeed, the metaphor of a drug "epidemic" is commonly used precisely because of that tendency for current users to "recruit" new users. If that were the only mechanism by which current use affected initiation, initiation might be expected to increase monotonically. Musto [12] has argued that, in addition, knowledge of the possible adverse effects of drug use deters or slows down initiation. He has hypothesized that drug epidemics eventually burn out when a new generation of potential users becomes aware of the dangers of drug abuse and, as a result, does not start to use drugs. Whereas many light users work, carry family responsibilities and generally do not manifest obvious adverse effects of drug use, a significant proportion of heavy users are visible reminders of the dangers of using addictive substances. Hence, it might be expected that large numbers of heavy users would suppress rates of initiation into drug use. It seems plausible that any reasonable model of an endogenous initiation might have the following properties:

(a) The rate at which current users recruit initiates is proportional to the number of light users. It is assumed that heavy users do not recruit initiates because they manifest ill effects of drug use and/or because they have been using drugs so long that they are older and socially distant from youth in the prime initiation ages;

(b) The rate at which current light users recruit initiates is moderated by the "reputation" or image the drug has and that reputation is governed by the relative number of heavy and light users, not the absolute number of heavy users. Even if there were a number of heavy users, the drug might appear benign if they were buried in a mass of (relatively happy) light users;

(c) Although most new users are recruited, for others the impetus to use drugs is internal. In the language of diffusion models [13], those individuals are "innovators" who initiate on their own for the sake of curiosity, by shifting from other drugs, or for some other reason, but not through the urging of someone who is already a user.

About 60 functional forms incorporating those features have been investigated, where Behrens and others [7] choose one of those five functional forms which give the best system performance with respect to minimization of the squared differences between modelled and observed initiation data from 1970 to 1991.

The rest of the so-called LH model [7] is essentially a continuous time analogue of Everingham and Rydell's model [10]. In that model, the population is divided into three groups: non-users, light users and heavy users (see the beginning of the present section). The number of non-users is assumed to be large enough compared with the number of users to behave like a constant and does not need to be modelled explicitly [14]. The flow rates from one state to another are assumed to be proportional to the source states and are computed as the time-continuous equivalents of the Everingham-Rydell estimates [15].

In addition to the endogenously modelled initiation, another difference between the model of Behrens and others [7] (illustrated in figure I) and that of Everingham and Rydell [10] is that, in the latter, the outflow from heavy use is divided into a flow out of use altogether (currently denoted by the "rate of desistance") and a flow back into light use. Behrens and others [7] dropped the latter flow for both theoretical and practical considerations. Theoretically, a flow from heavy to light use coupled with the Markov assumption implies that former heavy users who have de-escalated to light use and light users who had never been heavy users are indistinguishable. It is probably easier to relapse into heavy use than to enter the state for the first time, however. Hence, Behrens and others [7] prefer to have only a flow from heavy use to non-use and view that rate as net of relapse.

Before policy conclusions can be drawn from such a dynamic model of illicit drug use, it is necessary to make sure that the observed drug epidemic can be replicated by that model. (In the terminology of a mathematician, this is a "necessary" but not "sufficient" condition for further analyses.) As figure II shows, the fit of the modelled epidemic is not perfect; the historical data reflect a higher, sharper peak in light use. Nevertheless, the similarity is striking, given that the actual epidemic was subject to a varying set of drug control interventions over time that could be responsible for deviations from the model's uncontrolled path. Likewise, idiosyncratic historical events, such as Len Bias' death and the sharp increases in prices in late 1989, could account for some of the differences between historical and modelled data. And, of course, a perfect fit cannot be expected for a relatively simple model of a very complicated process such as the current cocaine epidemic in the United States.

In their extensive analyses, Behrens and others [7] found that omitting the feedback effects of prevalence on initiation was of relatively little consequence for

Figure I. Flow diagram for the LH model

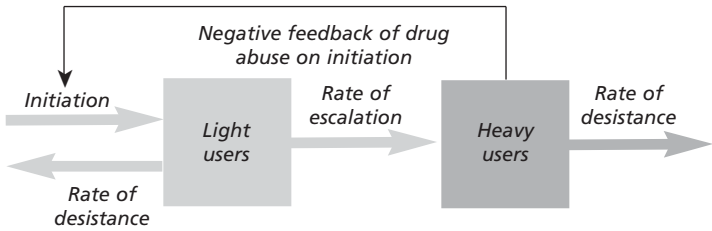
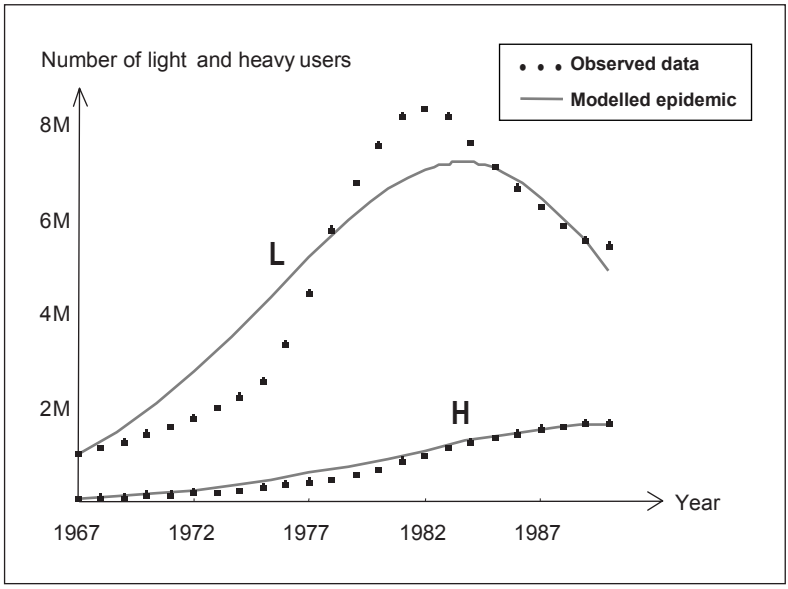


Figure II. Time paths of the continuously modelled cocaine epidemic in the United States of America and the smoothed historical data [10]



the analysis of the effectiveness of treatment and enforcement at a particular point in time, as Rydell and Everingham did [16]. It is of enormous consequence, however, for understanding how effective prevention programmes are or for understanding how the effectiveness of an intervention such as treatment might vary over the course of an epidemic. Although Behrens and others [7] did not investigate an optimal control model but a purely descriptive model, they did derive a number of interesting results with respect to the nature of drug control interventions by means of simple sensitivity analysis. For example, they found that different strategies were most effective at different stages of an epidemic, and one would expect the optimal mix of interventions to depend significantly on the course and status of the epidemic. More precisely, they hypothesized that prevention programmes might be most effective at early stages of the epidemic, when most users were light users, whereas treatment programmes might be most effective when a greater proportion of users are heavy users, as is typical for later stages of an epidemic.

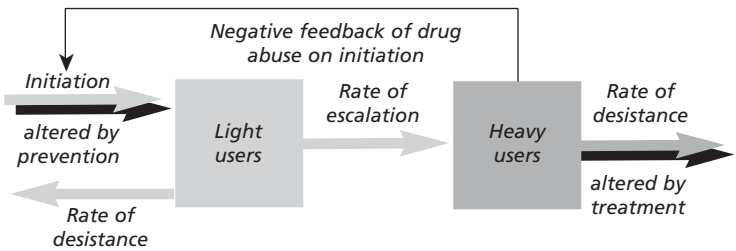
Inasmuch as it makes sense to vary the mix of interventions over the course of an epidemic, the authors intended to apply optimal control theory to their modelled epidemic—as a topic of further research in that area. Even though the extension seems to be the next logical step, it should be noticed that, generally, the derivation of the optimal choice of one or several controls in a dynamic model is a very complex and sophisticated task—also for models like the one created by Behrens and others [7], which in other respects are quite simple. The section below deals with that optimal control undertaking.

An optimal control model

From the descriptive model developed by Behrens and others [7], one may conclude that drug control interventions should change over time—especially over the course of a drug epidemic. The way they do so depends to a crucial extent on the choice of the interventions, on the objective and finally on the choice of restrictions on the drug control budget.

Behrens and others [9] formulated and solved an optimal control model to derive optimal intertemporal treatment and prevention spending decisions under three different assumptions with respect to restrictions on the drug control budget. In particular, the original model (presented in the previous section and illustrated in figure I) has been extended in that two of the flows are influenced ("controlled") by suitable control instruments: (primary) prevention decreases initiation by a certain percentage, while treatment of heavy users increases their rate of desistance, as illustrated in figure III. The objective chosen by Behrens and others [9] was to minimize the total social costs rather than to maximize social welfare in the sense that drug users' consumer surplus was excluded from the objective functional. The total social costs included both the social costs caused by illicit drug use and the additional monetary costs of the control measures (i.e. treatment and prevention spending). An alternative objective for the LH model has recently been presented by Kaya [17], namely, to reach some predetermined target in optimal time.*

Figure III. Flow diagram for the optimally controlled LH model



As mentioned above, Behrens and others [9] considered three different assumptions for restricting the drug control budget. These three cases can be described as follows:

(C1) The budget is constrained to be proportional to the size of the cocaine problem** and the proportions of that budget going towards treatment and prevention, respectively, are chosen once and fixed for all time;

*In particular, Kaya [17] considered the problem of finding a time-optimal control to get from some initial state (i.e. initial numbers of light and heavy users) to a target state.

**According to Rydell and others [1], total consumption is a reliable measure for the size of a drug problem.

(C2) The budget is chosen as in case (C1), but its allocation between treatment and prevention can be varied over time;

(C3) The budget is unconstrained in that both treatment and prevention spending can be chosen to be any non-negative number at all times. Theoretically, this case is the most reasonable one, because at some stages of the epidemic, high expenditures may be useful, while at other stages spending less money may be preferable. Practically, however, the implementation of the optimal solution to this problem may cause problems because the optimal expenditures can be considerably high or vary significantly over time.* It should also be noted that this model is more appropriate than one with a constrained budget if treatment and prevention resources are not allocated from a single source.

Comparing the results of these three constrained and unconstrained optimization models sheds light on how different forms of political constraints affect drug control. Insights of the optimally controlled LH model [9] include:

(a) Applying static interventions to a dynamic process may be counterproductive. This means that control measures, such as treatment and prevention, are most appropriate for specific stages of a drug epidemic and budget allocations across those measures should change over time. For instance, prevention works best when there are relatively few heavy users, that is, at the beginning of an epidemic. Treatment, on the other hand, is relatively more efficient at supporting the decline of drug abuse later in the epidemic (see figure IV);

(b) The transition period, when it is optimal to use both prevention and treatment extensively, is brief (see figure IV);

(c) Some control, even a "dumb" one in the case where not only is the budget constrained in total size, but also the shares of that budget being spent on treatment and prevention are chosen once and fixed for all time, does better than no control at all (see figure V);

(d) People who perceive drug use to be costly for society should favour greater drug control spending per gram consumed and allocate a greater share of that spending to prevention. Generally, it would be most effective to provide very large financial resources for control measures right from the onset of an epidemic (for prevention programmes), even if it might be difficult to justify doing so by the magnitude of the problem at that time;

(e) Total social costs increase dramatically if control is delayed (see figure VI).

---

*Tragler and others [18] derived the optimal solution for an alternative optimal control model of the United States cocaine epidemic. Their model is different in that they consider "average" users (i.e. there is no distinction between different levels of use) and the controls are treatment and price-raising enforcement. They show that, if initiation into drug use is an increasing function of the current number of users and control measures are implemented early in the drug epidemic, then it is optimal to use very large amounts of both enforcement and treatment to eradicate the epidemic. In other words, one would need a very large budget to pursue an optimal control of a drug epidemic that is still in its early stages, that is, small. In such a case, the per-user budget is enormous and the optimal policy will probably not be implemented because the public would not accept very large expenditures for a problem that is hardly visible.

Figure IV. Prevention and treatment spending for the unrestricted optimal control model (described in case (C3) above)

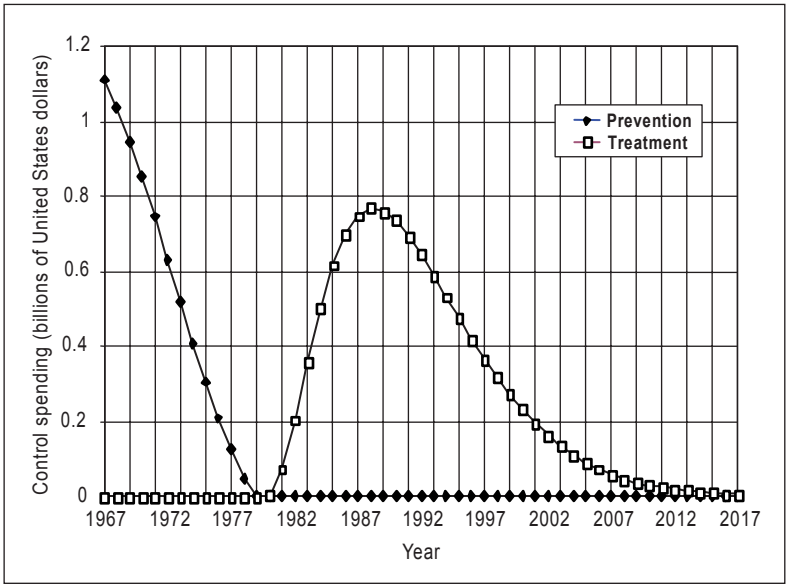


Figure V. Total quantity consumed during the current United States cocaine epidemic as well as controlled quantities for the different budget rules

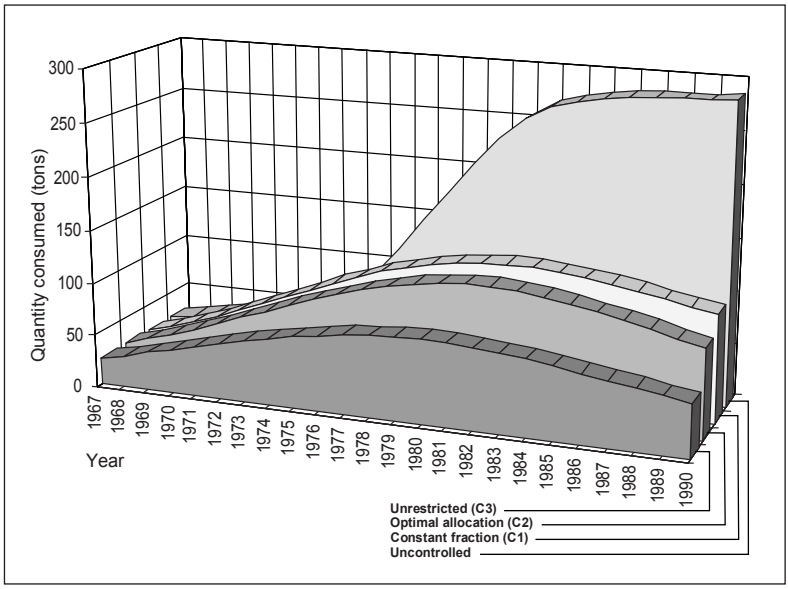
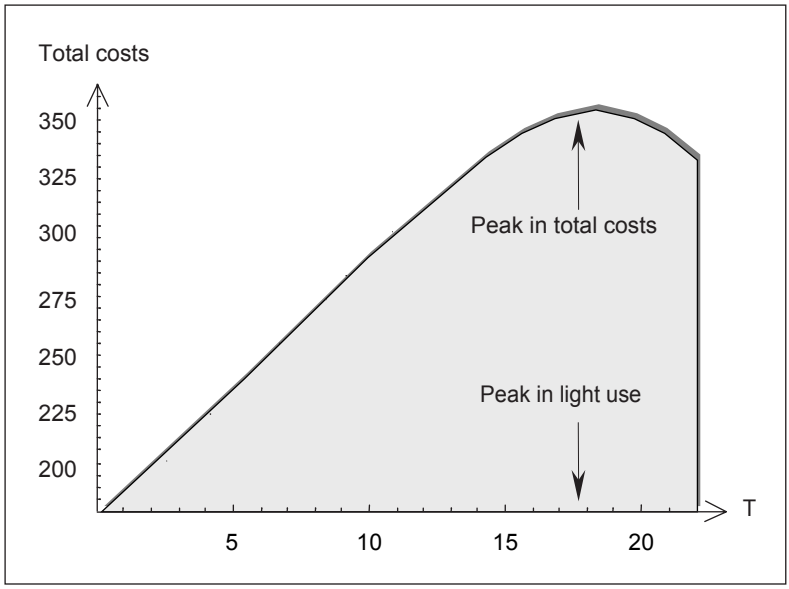


Figure VI. Total costs as a function of T, where T denotes the time when government starts to control (with T = 0 representing 1970 conditions) (Billions of United States dollars)



Looking at these insights from the study by Behrens and others [9] more carefully, it is evident that most of the results are as could have been expected (e.g. controlling the epidemic is good, delaying the control is bad, the controls should vary over the course of an epidemic and so on). On the other hand, the conclusions to be drawn from figure IV are to some extent awkward: it follows that it is optimal to stop prevention when the epidemic is still in its early stages, whereas treatment should not be implemented before the epidemic has "matured" somewhat. That result follows from the implicit model assumption that large numbers of heavy users are not only bad in the sense that they consume at high rates and hence impose large costs on society, but also good in the sense that they tend to discourage initiation. In other words, heavy users do impose costs in the near term, but they also generate a perverse sort of "benefit" for the future by reducing current initiation and thus future use. Since the timing of treatment and the reputation of the drug strongly interact in the framework of the model described above, the reputation mechanism deserves reconsideration. One possible refinement of the LH model can be obtained by assuming that the reputation, which influences initiation, is not a function of the current number of heavy users but rather of the memory of past heavy users; that framework is discussed in the next section.

Modelling a memory of drug abuse

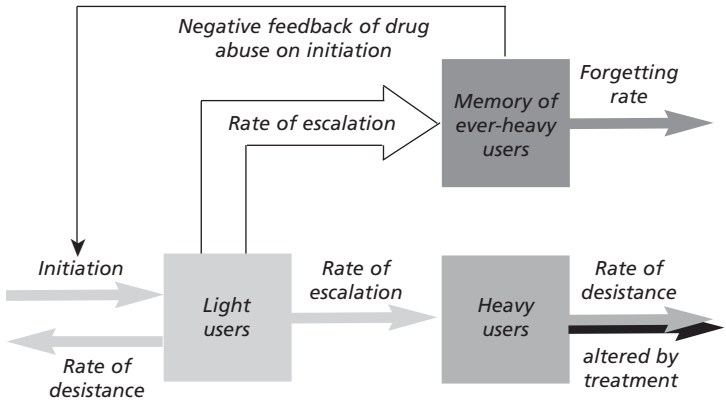
The extension of the reputation function to be a function of the decaying memory of heavy users carried out by Behrens and others [8] has obvious appeal. As

mentioned already, all knowledge of the negative experiences of a heavy user is unlikely to disappear the moment that the individual exits the heavy-user population, in particular if the exit is by death from drug use (as opposed to ceasing use or moving out of the area). In the LH models described above, removing a heavy user immediately erased all memory of that individual, so it sometimes appeared preferable to allow a person to suffer rather than to help that person to recover. In other words, the benefits of helping heavy users directly were outweighed by the cost of not being able to set an example with the help of their suffering. Such inhumane policies are most likely to disappear in a framework where past users can be remembered.

From a technical point of view, this refinement of the LH model requires a third category, E (the number of so-called "ever-heavy" users), so the analysis becomes more complicated (see figure VII for an illustration of the LHE model). This extension of the model with an additional category does not significantly improve the system performance in the descriptive (i.e. uncontrolled) case. That is, the time paths of the numbers of light and heavy users look very much like those for the original LH model (see figure II). However, in the optimally controlled LHE model, the memory removes the seemingly perverse results of the LH model that the presence of a heavy user can be so valuable as a deterrent that successfully treating such users actually increases consumption in the long run. In particular, treatment is no longer counterproductive unless one fails to keep up the memory of the adverse consequences of abuse in an adequate way [8].

An interesting question investigated in the paper by Behrens and others [8] is the fascinating interaction between a society's present-orientation, its ability to remember the past and the occurrence of cycles in the future.* They prove once again the old adage that "those who forget the past are condemned to repeat it".

Figure VII. Flow diagram of the optimally controlled LHE model



*Note that cycles (i.e. repeated drug epidemics) may also occur in the LH model presented above; for details, see Behrens and others [7].

More precisely, the greater the deterrent power of memories of drug abuse, the less likely society is to wind up with a chain of drug epidemics. Additionally, they verify that it can be desirable to relive past epidemics—at least for myopic decision makers. Or, to put it in simple terms, "for those who forget the past and over-value the present, it may be optimal to have their future recreate the past". Finally, it is shown that it is optimal to apply prevention throughout the epidemic because moderating the contagious aspect of initiation reduces the likelihood of cycles and instability.

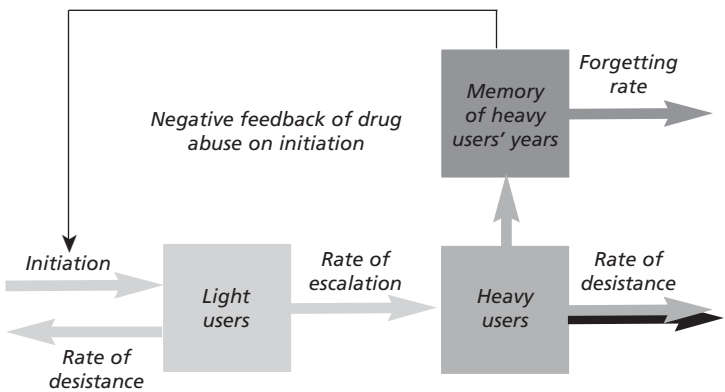
The results derived from the LHE model and their comparison with the conclusions of the LH study suggest that the extension to include a memory of people who have been ever-heavy users significantly improves the model performance. Nonetheless, the LHE model is not the ultimate model for including the implementation of a memory of drug abuse. The following section deals with another refinement of the model.

Refining the modelling of the memory of heavy use

The major drawback of the LHE approach is that three individuals who use drugs heavily for, say, one day, one year and one decade, would all contribute the same amount to the memory of heavy use. In reality, however, the longer an individual is addicted, the more problems he or she experiences, the greater the costs imposed on others, the more people there are who witness the behaviour and so on. So an appealing alternative is to base the negative reputation of an addictive substance not on the memory of the number of people who ever used drugs heavily, but instead on the memory of the number of heavy-user years, that is, the number of years spent in heavy use.

Behrens and others [19] have started to investigate such a model, where the number of ever-heavy users (E) is replaced by the number of heavy-user years (Y). That model, which is referred to as the LHY model, is illustrated in figure VIII.

Figure VIII. Flow diagram of the LHY model



From the preceding discussion it is clear that the LHE and the LHY models differ and that the LHY model provides a more realistic model formulation. A comparison of the respective flow diagrams (figures VII and VIII), however, does not reveal the fact that the analysis of the LHY model is also more difficult.*

What both models have in common, however, is the difficulty of estimating the parameters pertaining to the categories E and Y, respectively, for which there are no tangible quantities.** Nevertheless, at least for the LHY model, that problem has been resolved recently in a parameter estimation study with data on the current cocaine epidemic in the United States [20].

Behrens and others [19] make a thorough stability analysis of the uncontrolled LHY model's dynamics. The results obtained so far suggest, among other things, that drug prevention can temper drug prevalence and consumption and can avoid the reoccurrence of a drug epidemic. Furthermore, the results show the correlation between the deterrent power of negative experiences with drug abuse and the rate of forgetting them by providing a functional form for the phenomenon. The insights derived so far are general enough to allow a detailed characterization of what types of drugs—in terms of the probability of escalating to heavy use and the length of a typical addiction career—are most prone to generate cyclic, that is, reoccurring, drug epidemics. In addition, the LHY model even allows fairly general statements on epidemics of delinquent behaviour with a feedback effect of prevalence on initiation.

Summary and conclusions

Four dynamic models of the current cocaine epidemic in the United States have been reviewed in the present article. All differentiate between two levels of use ("light" and "heavy") and are based on Everingham and Rydell's 1994 model [10], but each has been developed by extending or refining another.

Firstly, the LH model by Behrens and others [7] is presented, which extends Everingham and Rydell's model [10] by introducing an endogenous function where light and heavy users feed back on initiation into light use ("infection" by light users versus "deterrence" by heavy users). As demonstrated in figure II, even such a "simple" model as the LH model is to some extent capable of reproducing such a "complicated" process as the current cocaine epidemic in the United States.

The optimally controlled LH model by Behrens and others [9] was then reviewed. The results derived from that study are interesting and most of them are not extremely surprising (e.g. controlling the epidemic is good, delaying the control is bad, the controls should vary over the course of an epidemic and so on).

*In contrast to the LHY model, the LHE model breaks into two parts (the LE and the H part) and, hence, the relevant dynamics for the descriptive case (but not for the optimally controlled case) can be transformed into a two-dimensional dynamical system, that is, the model may be investigated in the plane allowing a more thorough analysis, for example, with respect to cycles. This simplifies the analysis of the LHE model significantly.

**Obviously, this problem does not arise in the original LH models.

Still, one of the results (figure IV) suggested that the model needed some improvement: there are times in an epidemic when it is more desirable for a person to suffer than to help him or her, since the physical existence of a heavy user significantly diminishes initiation. Behrens and others [9] recognized that those seemingly "perverse" results were caused by the specific choice of the feedback function in the LH model, in particular, the assumption that heavy users would only contribute to a bad reputation of the drug as long as they were heavy users.

By introducing a third category (the "ever-heavy" users), Behrens and others [8] were able to formulate a refined model. One of the interesting questions investigated for that model is the fascinating interaction between a society's present-orientation, its ability to remember the past and the occurrence of cycles in the future.

In the LHE model, for the (negative) reputation of the drug it did not matter how long a heavy user was using drugs heavily. That problem was resolved by replacing the number of ever-heavy users with the number of heavy-user years. A short introduction to the LHY framework [19] was provided.

In conclusion, firstly, models such as the LHY model may be general enough to be applied to other drugs (both licit and illicit) and even to other forms of delinquent behaviour, provided they include a feedback effect of prevalence on initiation.

Secondly, it is clear from the studies reviewed here that dynamic models provide an important contribution to the problem of designing better drug control policies, because drug epidemics are, by definition, dynamic.

Thirdly, as has been shown here, the development of dynamic models of illicit drug consumption is itself a dynamic process, in which existing models are extended or refined. That improvement, however, depends to a crucial extent on improvement of the data sources, a better understanding of the underlying epidemic processes or new development of the tools needed for the analysis of dynamic models (both hard- and software).

References

1. C. P. Rydell, J. P. Caulkins and S. S. Everingham, "Enforcement or treatment: modeling the relative efficacy of alternatives for controlling cocaine", *Operations Research,* vol. 44, No. 6 (1996), pp. 687-695.
2. W. E. Schlenger, "A systems approach to drug user services", *Behavioral Science,* vol. 18, No. 2 (1973), pp. 137-147.
3. G. Levin, E. B. Roberts and G. B. Hirsch, *The Persistent Poppy: a Computer-Aided Search for Heroin Policy* (Cambridge, Massachusetts, Ballinger Publishing Company, 1975).
4. L. K. Gardiner and R. C. Schreckengost, "A system dynamics model for estimating heroin imports into the United States", *System Dynamics Review,* vol. 3, No. 1 (1987), pp. 8-27.

5. J. B. Homer, "A system dynamics model for cocaine prevalence estimation and trend projection", *Journal of Drug Issues,* vol. 23, No. 2 (1993), pp. 251-279.
6. J. B. Homer, "Projecting the impact of law enforcement on cocaine prevalence: a system dynamics approach", *Journal of Drug Issues,* vol. 23, No. 2 (1993), pp. 281-295.
7. D. A. Behrens, J. P. Caulkins, G. Tragler, J. L. Haunschmied and G. Feichtinger, "A dynamical model of drug initiation: implications for treatment and drug control", *Mathematical Biosciences,* vol. 159, 1999, pp. 1-20.
8. D. A. Behrens, J. P. Caulkins, G. Tragler and G. Feichtinger, "Why present-oriented societies undergo cycles of drug epidemics", *Journal of Economic Dynamics and Control,* forthcoming.
9. D. A. Behrens, J. P. Caulkins, G. Tragler and G. Feichtinger, "Optimal control of drug epidemics: prevent and treat—but not at the same time?", *Management Science,* vol. 46, 2000, pp. 333-347.
10. S. S. Everingham and C. P. Rydell, *Modeling the Demand for Cocaine* (Santa Monica, California, RAND Corporation, 1994).
11. National Institute on Drug Abuse, "Overview of the 1991 household survey of drug abuse", press release, December 1991.
12. D. F. Musto, *The American Disease: Origins of Narcotic Control* (New York, Oxford University Press, 1987).
13. J. P. Gould, "Diffusion process and optimal advertising policy", *Microeconomic Foundations of Employment and Inflation Theory,* E. S. Phelps and others, eds. (London, Macmillan, 1970), pp. 338-368.
14. S. S. Everingham, C. P. Rydell and J. P. Caulkins, "Cocaine consumption in the US: estimating past trends and future scenarios", *Socio-Economic Planning Sciences,* vol. 29, No. 4 (1995), pp. 305-314.
15. S. S. Everingham and C. P. Rydell, op. cit., 1994, p. 43.
16. C. P. Rydell and S. S. Everingham, *Controlling Cocaine: Supply Versus Demand Programs* (Santa Monica, California, RAND Corporation, 1994).
17. C. Y. Kaya, "Time-optimal switching control for the US cocaine epidemic", paper presented at the Workshop on Dynamic Drug Policy: Understanding and Controlling Drug Epidemics, Vienna, 22-24 May 2000.
18. G. Tragler, J. P. Caulkins and G. Feichtinger, "Optimal dynamic allocation of treatment and enforcement in illicit drug control", *Operations Research,* vol. 49, No. 3 (May-June 2001), pp. 352-362.
19. D. A. Behrens, J. P. Caulkins, G. Tragler and G. Feichtinger, "Memory, contagion, and capture rates: characterizing the types of addictive behavior that are prone to repeated epidemics", discussion paper of the College of Business Administration, University of Klagenfurt, 2000.
20. C. Knoll and D. Zuba, "LHY-model", *Students' Project,* Department of Operations Research and Systems Theory, Vienna University of Technology, 2000.

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

Pharmaco-economics of drug addiction: estimating the costs of hepatitis C virus, hepatitis B virus and human immunodeficiency virus infection among injecting drug users in member States of the European Union*

M. J. POSTMA
Groningen University Institute for Drug Exploration/Groningen Research Institute of Pharmacy, Groningen, the Netherlands

L. G. WIESSING
European Monitoring Centre for Drugs and Drug Addiction, Lisbon, Portugal

J. C. JAGER
National Institute of Public Health and the Environment, Bilthoven, the Netherlands

ABSTRACT

The objective of the research described in the present article was to estimate the costs of hepatitis C virus (HCV), hepatitis B virus (HBV) and human immunodeficiency virus (HIV) infection by intravenous drug use in countries of the European Union.

The pharmaco-economic model applied linked incidence to lifetime cost estimates (incidence-based model). Estimates are presented within a broader framework of social costs.

In the baseline, estimated costs of HCV, HBV and HIV infection related to drug addiction amount to 1.89 euros (€). In sensitivity analysis, costs up to €2.57 are estimated (0.5 per cent of total European Union expenditures on health care).

A preliminary estimate for the costs of HCV, HBV and HIV infection related to drug addiction in the European Union is provided. Those costs reflect a relevant share of total costs related to drug addiction.

Introduction

Pharmaco-economics covers the research field of allocating scarce resources to drugs and drug-related activities [1, 2]. The primary field of pharmaco-economical

*The authors acknowledge the support of Fernando Antoñanzas (University of La Rioja, Logroño, Spain), Ardine de Wit (National Institute of Public Health and the Environment, Bilthoven, the Netherlands) and Eric Single (University of Toronto, Canada).

applications covers the medical applications of drugs, with relevant areas such as costing studies and cost-benefit analyses of pharmaceuticals. The methodology can however be directly applied to "unhealthy" applications of drugs as well. The scope of the present study is limited to frequent smoking and injecting drugs such as opiates, benzodiazepines, amphetamines and cocaine, referred to here in brief as "drug addiction".

Pharmaco-economics has been applied to drug addiction previously in only a few studies. For example, Kim and colleagues analysed costs and benefits of drug prevention programmes for the United States of America over the period 1979-1992 [3]. They estimated a favourable cost/benefit ratio: there is a saving in direct and indirect costs of $15 for each $1 spent in drug addiction prevention. Rice and colleagues estimated the costs of drug addiction-related illnesses within a broader social costing framework [4]. That framework included both the indirect costs of production losses and the costs of crime.

As regards drug addiction, the pharmaco-economic concept of social costs provides a very suitable approach. Social costs reflect the total burden of costs for society. The societal perspective is much wider than the health-care perspective, which is limited to direct health-care costs. Since it includes indirect costs, the societal perspective is often the viewpoint that policy makers are most interested in, as such evaluations assist in deciding on the allocation of resources between different sectors and generic methods developed allow valid comparisons to be made between very different programmes [5]. One field enjoying a growing interest involves the pharmaco-economics of prevention of infectious diseases. For example, in the Netherlands, investigations into the pharmaco-economics of hepatitis B virus (HBV) vaccination and human immunodeficiency virus (HIV) screening programmes are explicitly required by the Ministry of Health before implementation of a programme [6]. Pharmaco-economic evaluation of drug addiction is closely related to prevention of infectious diseases, as transmission of bloodborne diseases is among the most serious health and economic consequences of injecting drug use. The authors of the present article consider the costs of drug-related acquisition of hepatitis C virus (HCV), HBV and HIV in member States of the European Union (EU).

A methodology derived from pharmaco-economic theory to classify the costs related to drug addiction and to provide partial estimates is presented below. The classification draws on general economic principles of costing and the estimation procedure has previously been applied to costs related to HIV/acquired immune deficiency syndrome (HIV/AIDS) at the EU level [7]. The estimation is based on the direct medical costs of HIV/AIDS and hepatitis epidemics among drug users in EU. The preliminary estimate provided only covers part of all social costs for diseases related to drug addition. Indirect costs of the infectious diseases and the costs of other (non-infectious) diseases are not considered here. Furthermore, costs of addiction treatment, criminal acts and low performance in jobs and indirect costs of mortality due to overdoses are also left for further research. However, a crude estimate is given here for the costing of infectious diseases in EU through drug addiction that has not previously been attempted.

The pharmaco-economic model

The authors' pharmaco-economic model was originally applied to 10 EU countries (Ireland and Luxembourg were left out because of lack of data). In that model, epidemiological information (incidence of infections) and economics (lifetime costs per infection) were linked to estimate the direct medical costs of drug-related HBV, HCV and HIV.

Traditionally, pharmaco-economic research uses two basic methods to express the impact of drugs and disease, known as incidence-based and prevalence-based methods. The first involves accounting lifetime costs to the year of incidence of disease, while the second method involves a breakdown of those lifetime costs over the relevant stages of disease into stage-specific prevalence estimates with a view to, for example, annual costing estimates for budgetary purposes [7]. To cover the full pharmaco-economic impact of drug addiction, it is necessary to analyse the lifetime cost consequences incurred by a representative incident of infection of a drug user. For direct medical costs, this incidence-based approach does however assume that future health-care technologies will remain the same as in the analysis. In addition, all the other assumptions are extended up to the time when the last incident of infection of drug users has been dealt with.

If the assumption of everything staying equal holds true for the period of analysis and the full pharmaco-economic impact is what is being referred to in a decision context, the incidence-based approach is the more appropriate one. This could be the case if the decision context, for example, refers to investments and allocating budgets to prevent problems related to drug addiction. Therefore results using the incidence-based approach are reported here.

Data and assumptions

Epidemiology

Epidemiological information refers to the mid-1990s (the most recent published information available). HIV incidence among drug users in EU is estimated by back calculation. Estimates correspond to a range from approximately 250 per million population in Spain and Portugal to less than 10 per million in Germany, Greece, the Netherlands and the United Kingdom of Great Britain and Northern Ireland [8].

Incidence of drug-related hepatitis was estimated using a review of European prevalence studies for HCV, HBV and HIV in specific groups of drug users (see table 1) [9]. Excess prevalence rates for HCV and HBV were applied over the HIV prevalence rate to HIV incidence (for example, if the HBV antibody is twice as prevalent as that of HIV in a sample of Spanish drug users, Spanish HBV incidence was assumed to be twice that of Spanish HIV).*

*This straightforward application of prevalence relationships to incidence certainly neglects differences in hepatitis and HIV mortality rates, potentially disturbing similar patterns in prevalence and incidence. HIV prevalence estimates were available for all countries considered. For Portugal, in the absence of an estimate for HBV antibody prevalence, an assumption of 50 per cent of HCV antibody prevalence was made, as this crudely reflects the pattern of HCV/HBV prevalence in the other countries considered.

Table 1. Prevalence rates of infections with hepatitis C virus, hepatitis B virus and human immunodeficiency virus infection applied in the pharmaco-economic model[a]

	Prevalence rate (percentage)		
	Hepatitis C virus	*Hepatitis B virus*	*Human immunodeficiency virus*
Belgium	47	22	1.4
Denmark	50	21	4.0
France	66	23	18.0
Germany	79	64	3.9
Greece	65	65	1.3
Italy	67	40	16.0
Netherlands	79	61	14.0
Portugal	74	37	14.0
Spain	83	59	30.0
United Kingdom of Great Britain and Northern Ireland	38	19	0.6

[a]Adapted from L. G. Wiessing and others, "Impact and control of AIDS, HIV, and hepatitis B and C among injecting drug users in Europe: an exploratory overview", Inaugural Meeting Report by National Institute on Drug Abuse, National Institutes of Health, United States Department of Health and Human Services, Geneva, 25 and 26 June 1998.

Health care

Hospital bed needs arising in relation to HIV/AIDS infection vary over the different stages of the disease. Furthermore, hospital costs and the costs of pharmaceuticals are the major components of health-care costs to combat HIV/AIDS. In particular, the terminal phase of AIDS is accompanied by a relatively high level of hospital bed needs. In this regard, two stages were previously differentiated for AIDS: a final stage of maximally six months before death (the late stage (LS)) and a stage for the foregoing period (the chronic stage (CS)). Application of the stage concept to France, Greece, Italy, the Netherlands, Spain and the United Kingdom has shown that per person-year inpatient days in LS are at least twice those in CS. Inpatient day needs in symptomatic pre-AIDS stages equal 20 per cent of those in CS. Information on inpatient days was available for the six countries mentioned [7].

To allow for recent advances in pharmaco-therapy, the previously published progression-of-disease model was extended with a stage reflecting the life-years gained on highly active anti-retroviral therapy (HAART), with hospital inpatient days needs similar to those in the symptomatic pre-AIDS stages [10]. To summarize the full progression-of-disease model for HIV: on average six years of asymptomatic HIV infection, one extra year gained on HAART, six years of symptomatic pre-AIDS disease and two years of AIDS (CS and LS) [7, 10].

Costing

Costs for one year on HAART were assumed to total 10,000 euros (€) for the package of pharmaceuticals (for example, zidovudine, lamivudine and ritonavir). Country-specific costs per acute care day were used, varying from €420 (United Kingdom) to €600 (France).* Costs per acute care day include average outpatient costs. Outpatient costs were corrected for by lowering costs per acute care day by 30 per cent in the baseline (based on Netherlands data [7]) and 10 per cent in sensitivity analysis. Costs per outpatient contact were assumed to be 25 per cent of corrected unit costs per acute care inpatient day [7].

Monetary costs were discounted for time preference at 4 per cent [12], 1 per cent and 7 per cent [13] in sensitivity analysis. Finally hospital costs for HIV/AIDS were assumed to reflect 90 per cent of total costs, excluding the costs of HAART [7]. As detailed information on health care for HIV was only available for six countries (France, Greece, Italy, the Netherlands, Spain and the United Kingdom), for the other countries the cost figures of neighbouring countries were applied (that is, the Netherlands estimate for Belgium, Denmark and Germany; the Spanish figure for Portugal).

Only crude lifetime cost estimates for incidence of HCV and HBV infection could be used as very little information was available [14-17]. That information suggests that average lifetime costs per HCV infection are in the order of magnitude of 10-20 per cent of lifetime HIV costs and that lifetime HBV costs are 10 per cent of lifetime HCV costs [18, 19]. In the baseline it was assumed that lifetime HCV costs were 15 per cent of lifetime HIV costs; in two alternative scenarios the lower and upper limits of 10 and 20 per cent, respectively, were investigated.

Results and discussion

Figure I shows estimated lifetime costs of HIV infection for the six countries for which detailed information was available. These varied from €42,500 for the United Kingdom to €90,800 for France.*

In the baseline, estimated costs of drug-related HCV, HBV and HIV infection amount to €1.89 billion for EU as a whole. HCV infection accounts for approximately 40 per cent of those costs (see figure II), HIV infection for the major proportion of the costs. Figure III indicates the distribution over the 10 countries considered: Spain and Italy are estimated to account for a major proportion of total EU costs (approximately €1.4 billion). Both countries show relatively large shares for drug-related HIV infection costs. Furthermore, results indicate that in the United Kingdom and Germany costs arise primarily from hepatitis.

*Based on purchasing power parities for gross domestic product as estimated by the Organisation for Economic Cooperation and Development [11]. For the two countries with missing cost data, Greece and Italy, €500 was assumed.

Figure I. Estimated lifetime costs of human immunodeficiency virus infection in six member States of the European Union

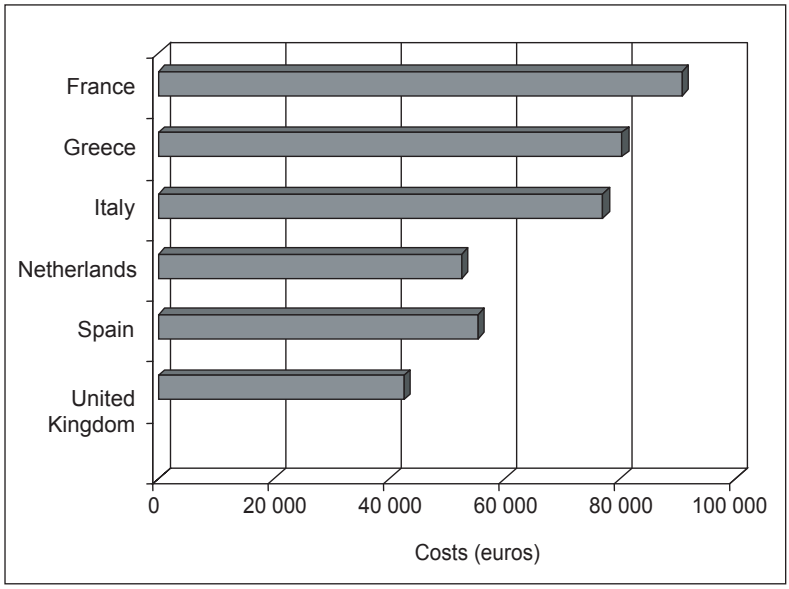


Figure II. Distribution of estimated drug-related costs for hepatitis C virus, hepatitis B virus and human immunodeficiency virus infection for the European Union as a whole, excluding Ireland and Luxembourg

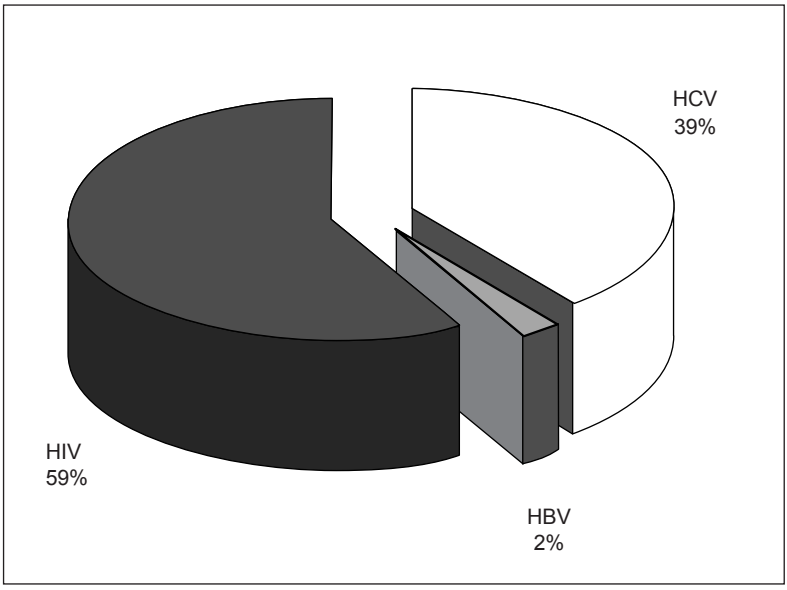
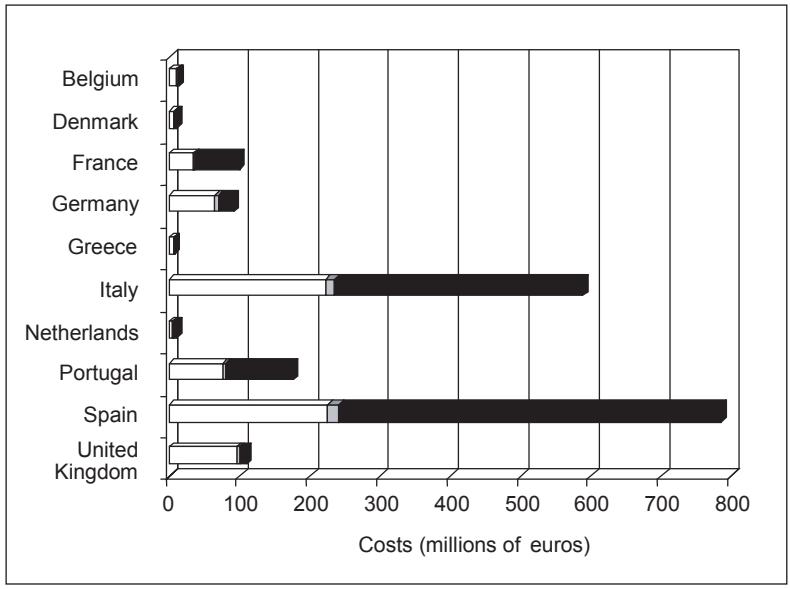


Figure III. Costs of hepatitis C virus, hepatitis B virus and human immunodeficiency virus infection for 10 member States of the European Union[a, b]



[a]White = HCV, Grey = HBV, Black = HIV

[b]These country-specific results can only be considered indicative; differences between countries depend to a crucial extent on the assumptions derived from the European prevalence studies (see table 1).

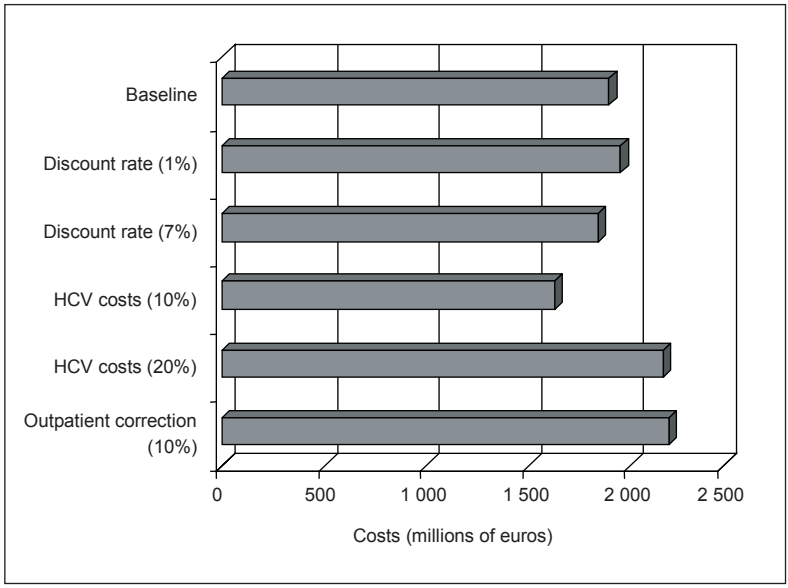
Figure IV shows the results of one-way sensitivity analysis. Variation in the discount rate induces the smallest changes in total costs. Estimated costs range from €1,630 to €2,159 (minus 14 per cent and plus 16 per cent of the baseline, respectively). Multi-way sensitivity analysis indicates estimated costs up to €2,570 million (HCV costs at 20 per cent of HIV costs, 10 per cent lowering of acute care costs for outpatient costs and 1 per cent discount rate).

The types of estimate described above may serve several purposes. They can, for example, be used in budget analysis. Estimates of the costs of drug addiction are typically related to health-care budgets, at the national and EU levels. In 1995, the estimated costs of drug-related HCV, HBV and HIV infection as a percentage of expenditure on health in 1995 were as follows (in euros at purchasing power parities for gross domestic product):*

Belgium	0.09
Denmark	0.12
France	0.09
Germany	0.06
Greece	0.08

*Source: Organisation for Economic Cooperation and Development, *Health Data 99* (Paris, 1999).

Figure IV. Sensitivity analysis for total costs of drug-related hepatitis C virus, hepatitis B virus and human immunodeficiency virus infection for the European Union as a whole, excluding Ireland and Luxembourg[a]



[a]Baseline, variation in the discount rate (4 per cent in the baseline), variation in the level of lifetime HCV costs as a percentage of lifetime HIV costs and correction for outpatient costs in costs of acute care day by 10 per cent (20 per cent in the baseline; see text).

Italy	0.73
Netherlands	0.04
Portugal	1.86
Spain	2.07
United Kingdom of Great Britain and Northern Ireland	0.16

Relatively high shares are estimated for the southern European countries Italy, Spain and Portugal, with significant impacts of HIV. As for northern European countries, Germany and the United Kingdom have relatively high shares, estimated to be caused primarily by HCV. The baseline estimate for drug-related costs of HCV, HBV and HIV amounts to approximately 0.4 per cent of expenditure on health in EU as a whole (up to 0.5 per cent in sensitivity analysis).

Two other studies provide a framework to compare with the estimate presented here (table 2) [20, 21]. Both studies distinguish between direct health-care costs, other direct costs such as government spending and indirect costs of production losses. The table illustrates that the authors' estimate of the most relevant share of direct health-care costs related to HCV, HBV and HIV constitutes a significant contribution to a total estimate of social costs. However, large shares also remain open to further research. In particular, further research should consider

indirect costs at the EU level. In that respect the distinction between the human-capital method and friction-costing method for indirect costing is relevant [22]. The latter is favoured in some European countries and significantly reduces indirect costs related to mortality compared with human-capital costing. Application of the friction-costing concept would raise the share of direct health-care costs in the study by Single and colleagues [21] from the current 6 per cent in the table to 10-12 per cent.

Table 2. Distribution of estimated social costs of drug addiction in Canada [21] and France [20], various cost categories *(Percentage)*

Costs	*Canada*	*France*
Direct health care[a]	6	12
Direct other[b]	34	36
Indirect[c]	60	52

[a]Costs of HIV/AIDS and Subutex treatment only.

[b]Comprises costs for Governments and costs of prevention, research and law enforcement.

[c]Indirect costs for absence due to sickness and for mortality, using the human-capital method of enumeration.

The present authors' results depend to a crucial extent on the epidemiological information used in the model. Estimated incidences reflect the situation in the mid-1990s and were based on modelling (HIV) and relative prevalences of HCV and HBV in samples. For example, given that in the United Kingdom samples considered HCV is approximately 100 times more prevalent than HIV, an estimated HCV incidence for the United Kingdom of more than 100 times HIV incidence results. The epidemiological information used is the only information available at the moment, but the representativity of the samples for the national situations are subject to question. Next to, for example, the setting of the sample, one criterion of representativity concerns the size of the samples: sample sizes differ from 140 in Greece to 67,000 in Italy (the United Kingdom is in between, with approximately 2,000 drug users in treatment centres).

The incidence-based approach has been described in detail here. Cost estimates derived with the incidence-based approach are the most suitable estimates for use in cost-effectiveness analysis. [13] If, for example, the aim is to evaluate the cost-effectiveness of needle exchange or methadone programmes with respect to infections averted, lifetime costs related to one incident case of infection represent the most appropriate type of information. If, on the other hand, the budget impact of drug addiction is the focus of analysis and if that impact is expected to change with new technologies, the prevalence-based approach—applied over a well-defined period of analysis—is more appropriate to support decisions from the pharmaco-economic point of view. However, with respect to the long-term social costing or full pharmaco-economic impact, incidence- and prevalence-based approaches give the same result in the end.

Conclusion

In conclusion, a preliminary cost estimate for incidence-based drug addiction-related costs for HCV, HBV and HIV infection in EU member States has been provided. Further development of the model will be directed to designing a prevalence-based approach and estimating other categories of social costs.

References

1. J. L. Bootman, R. J. Townsend and W. F. McGhan, *Principles of Pharmacoeconomics* (Cincinatti, Ohio, Harvey Whitney Books, 1996).
2. European Monitoring Centre for Drugs and Drug Addiction, *Annual Report on the State of the Drugs Problem in the European Union* (Lisbon, 1997).
3. S. Kim and others, "Benefit cost analysis of drug abuse prevention programs: a macroscopic approach", *Drug Education,* vol. 25, 1995, pp. 111-127.
4. D. Rice, S. Kelman and L. Miller, "Estimates of economic costs of alcohol and drug abuse and mental illness, 1985-88", *Public Health Reports,* vol. 106, 1999, pp. 280-292.
5. M. F. Drummond and others, *Methods for the Economic Evaluation of Health Care Programmes* (Oxford, Oxford University Press, 1999).
6. M. J. Postma and others, "Opportunistic screening for genital infections with *Chlamydia trachomatis* among the sexually active population of Amsterdam: cost-effectiveness of screening women" (in Dutch), *Nederlands Tijdschrift voor Geneeskunde,* vol. 143, 1999, pp. 677-681.
7. M. J. Postma, "Assessment of the economic impact of AIDS at national and multinational level", thesis, Maastricht University, 1998.
8. A. M. Downs and others, "Reconstruction and prediction of the HIV/AIDS epidemic among adults in the European Union and in the low prevalence countries of central and eastern Europe", *AIDS,* vol. 11, 1997, pp. 649-662.
9. L. G. Wiessing and others, "Impact and control of AIDS, HIV, and hepatitis B and C among injecting drug users in Europe: an exploratory overview", Inaugural Meeting Report by National Institute on Drug Abuse, National Institutes of Health, United States Department of Health and Human Services, Geneva, 25 and 26 June 1998.
10. M. J. Postma, J. Van Bergen and J. C. Jager, "Pharmaco-economic evaluation of antiretrovirals in treatment and prevention of HIV/AIDS", *Clinical Science,* part II, proceedings of the 12th World AIDS Conference (1998), pp. 381-385.
11. Organisation for Economic Cooperation and Development, *Health Data 99* (Paris, 1999).
12. Netherlands Ministry of Finance, *Financial Information and Administration for the Government* (in Dutch) (The Hague, 1996).
13. M. E. Gold and others, *Cost-Effectiveness in Health and Medicine* (New York/Oxford, Oxford University Press, 1996).
14. G. A. De Wit and R. A. Welte, *Economic Evaluation of Hepatitis B Vaccination Strategies: a Systematic Review of the Literature* (Bilthoven, National Institute of Public Health and the Environment of the Netherlands (RIVM), 1999).

15. G. A. De Wit and others, "Cost-effectiveness of national hepatitis B vaccination"; interim report (Bilthoven, National Institute of Public Health and the Environment of the Netherlands (RIVM), 2000).

16. K. Brown and N. Crofts, "Health-care costs of a continuing epidemic of hepatitis C virus infection among injecting drug users", *Australia New Zealand Journal of Public Health,* vol. 22, 1998, pp. 384-388.

17. Personal communication from Ardine de Wit, National Institute of Public Health and the Environment of the Netherlands (RIVM).

18. S. Petrou and others, "The economic costs of caring for people with HIV infection and AIDS in England and Wales", *PharmacoEconomics,* vol. 9, 1996, pp. 332-340.

19. D. R. Holtgrave and S. D. Pinkerton, "Updates of cost of illness and quality of life estimates for use in economic evaluations of HIV prevention programs", *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology,* vol. 16, 1997, pp. 54-62.

20. P. Kopp, "Calculating the social cost of illicit drugs", interim report for the Council of Europe, Paris, 2000.

21. E. Single and others, "The economic costs of alcohol, tobacco and illicit drugs in Canada, 1992", *Addiction,* vol. 93, 1998, pp. 991-1006.

22. M. A. Koopmanschap and others, "The friction cost method for estimating indirect costs of disease", *Journal of Health Economics,* vol. 14, 1995, pp. 171-189.

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

Controlling infectious diseases among injection drug users: learning (the right) lessons from acquired immunodeficiency syndrome (AIDS)

H. POLLACK

School of Public Health, University of Michigan, Ann Arbor, Michigan, United States of America

ABSTRACT

For decades, infectious diseases have posed a serious and avoidable threat to the health and survival of injection drug users (IDUs). The most deadly threat currently arises from the human immunodeficiency virus (HIV) and the acquired immunodeficiency syndrome (AIDS). Yet less visible infectious diseases are also significant. Data from local and population-based surveys indicate that the great majority of IDUs in the United States of America are infected with hepatitis B or hepatitis C virus [1, 2]. A smaller but still significant proportion is infected with endocarditis, bone and joint infections [3], tuberculosis [4] and other infectious ailments. As the average age of IDUs increases, the prevalence of infectious disease may increase [3].

What can be done to address those threats? Can epidemiological models play a useful role in public health prevention efforts? Can treatment for substance abuse slow the spread of disease, and is treatment cost-effective when compared with other public health efforts? Can hepatitus C virus or other highly infectious diseases be controlled within drug-using populations? How can clinicians and policy makers target IDUs at greatest risk of getting and spreading blood-borne disease? The present paper offers five tentative principles to help answer those difficult questions.

Models matter

Many clinicians and policy makers are skeptical about the merits of formal models of the spread of infectious disease. Some skepticism is warranted. Injection drug users (IDUs) are a hidden population whose risk behaviours, and even whose absolute numbers, are only imperfectly known [5] Many basic parameters, such as the probability of infection from a single exposure, are poorly understood and differ across the drug-using population. The baseline rate of new infections is difficult to estimate within many high-risk groups. Rigorous evaluations of prevention are rare. When such evaluations do exist, they may not be applicable to local conditions. Even when sound epidemiological and programme evaluation data are available, this information may not be easily translated into politically and administratively feasible policy guidance.

Thus, the first question any policy analyst confronts may well be the hardest: Why bother? Although that question may be answered in many ways, the simplest and best answer is that models reveal and scrutinize the implicit assumptions already being used to fight infectious disease [6]. Rules of thumb often reflect unexamined assumptions that turn out to be implausible [7].

Kaplan and Pollack reviewed procedures used to allocate resources for human immunodeficiency virus (HIV) prevention in the United States [8-9]. Many United States policy makers try to allocate resources on the basis of the number of individuals in each risk group. That method of allocation is sensible when individual risk groups are similar and when the available prevention procedures are similarly effective across risk groups. When those assumptions go wrong, population-based resource allocation fails to provide an appropriate means of targeting resources to those in greatest need. Moreover, such allocation fails to prevent the largest number of new infections, given the resources available for interventions aimed at prevention [10]. In the real world, the incidence of HIV and the effectiveness of available programmes across risk groups vary greatly.

Worse, the political and organizational realities of group decision-making foster arbitrary decision-making. Altman, Greene and Sapolsky note that health planners respond to technical and political uncertainty by seeking "convenient proxies for need to be applied in allocation decisions"[11]. Wary of debating the merits of specific facilities, health system planners often resort to drawing up need-assessment formulas to evaluate proposed capital-intensive services. Economists have never doubted that such formulas provide poor guidance in evaluating the impact or cost-effectiveness of proposed expenditures. However, such methods find wide appeal as planners seek credible focal points to resolve internal disputes and justify controversial policies. For many of the same reasons, Kaplan and Pollack found the widespread application of complex and analytically ungrounded formulas to allocate funds for HIV prevention.

Models also help policy makers understand the linkage between readily observed data and the less visible, underlying pattern of the spread of disease. In facilitating the process of understanding, explicit models indicate which data have the most important influence on policy and how such data might be interpreted.

Evaluations of syringe exchange programmes (SEPs) illustrate the importance of such models. Although SEPs have not been scrutinized through prospective randomized trials, an impressive literature documents the effectiveness of such interventions [12-17]. A United States Secretary of Health and Human Services recently acknowledged that point, although she was not able to lift the current ban on United States federal funding [18, 19]. Although many studies have examined the impact of SEPs on self-reported behaviour change, such studies did not examine the relationship between SEPs and the rate of new HIV infections. Kaplan and Heimer addressed this critical policy concern by developing an elegant mathematical model to interpret HIV test data from exchanged needles [17].

Unfortunately, transmission modelling also indicates that the success of SEPs in curbing HIV transmission may not be replicated for other diseases. Short-term reduction in HIV transmission is sufficient to reduce long-run incidence and

prevalence because HIV, though deadly, is difficult to transmit. The analytic work of the author of the present paper and existing programme data suggest that the same programme performance is not enough to prevent the spread of more efficiently transmitted infectious diseases [20-22].

Apart from HIV, hepatitus C virus (HCV) is the most serious blood-borne epidemic among IDUs. So far, no intervention for prevention or treatment has been consistently found to slow the spread of the disease. Most discouragingly, many IDU populations display endemic HCV prevalence despite well-implemented SEPs and methadone maintenance treatment that successfully contain HIV [23-25]. Policy analysis has not reached a stage where it can offer reliable programme recommendations. In contrast to HIV, for which successful targeting of high-risk individuals may offer the greatest long-term benefit, the most effective prevention strategies for HCV might include more substantial education and outreach directed at low-risk IDUs.

Data matter

For analytic models to be of use, accurate and pertinent information is needed concerning both the spread of infectious disease and the performance of available interventions.

One of the most important but difficult tasks is to obtain usable (if imperfect) estimates of disease incidence, that is, the number and overall pattern of new infections in the drug-using population. Those estimates are important, because existing prevention measures are often based upon current disease prevalence, that is, the number of IDUs known to be infected. Ignoring many complexities, researchers can easily measure prevalence when infected individuals seek medical attention. In many settings, disease prevalence is estimated through the use of existing clinical data systems at hospitals, public clinics and other sites where health care is provided.

Unfortunately, this method of resource allocation is often inappropriate. Current prevalence indicates the past history of the spread of disease. In a dynamic blood-borne epidemic, prevalence estimates can provide a misleading guide to the specific risk-groups that are currently experiencing high rates of new infection. In the area of HCV, incidence analysis indicates that young and inexperienced IDUs are experiencing remarkably high rates of new infection [1, 22].

Many policy makers in public health are well aware that disease incidence is the touchstone of prevention efforts. The practical challenge is to estimate disease incidence in a hidden population frequently estranged from the health-care delivery system. Both technology and improved public management can play an essential role in meeting this challenge. In the area of HIV prevention, new statistical and chemical testing methods have become available to enable the rate of new infection in specific populations to be directly estimated [26, 27]. Many of these methods are easily transferred to other blood-borne ailments. Other advanced techniques such as molecular tracing are also helpful in tracing outbreaks and documenting novel modes of transmission.

Apart from technological innovation, the most important challenge is to treat the surveillance of infectious disease as a central tool of public health policy rather than as the stepchild of existing convenient data systems [10]. Because many IDUs have limited contact with the health-care delivery system or with public health services, incidence and prevalence estimates may be biased if they are drawn from HIV/AIDS treatment, voluntary testing and other services that reach small, often self-selected, populations. Such data must be supplemented by new sources that include more hidden populations of IDUs.

Careful sampling of IDUs in other contexts is therefore essential to provide an accurate picture of disease incidence and prevalence in the drug-using populations. Drug-treatment clients and SEP clients are critical populations that bear careful study in population-based surveillance. Out-of-treatment IDUs may also be identified through epidemiological studies of prison inmates [28, 29]. Perhaps most important is to explore the rate of new infections among out-of-treatment IDUs who do not use other services [30]. Many of them are probationers or are under other forms of judicial supervision [31].

Substance-abuse treatment should be used to prevent new infections

Treatment for substance abuse plays a critical role in improving the well-being and social performance of drug users. Many studies document the effectiveness, and cost-effectiveness, of methadone maintenance and other forms of treatment in reducing criminal activity, improving health and increasing employment among treatment clients [32-36].

The value of treatment was underscored by a randomized study of methadone maintenance among Swedish IDUs [32, 37]. Within two years of the inception of the study, 2 out of 17 members of the non-methadone control group died from apparent overdose. One other member of the non-methadone group suffered a leg amputation, while two others suffered severe infection. Two of the remaining members of the non-methadone group were incarcerated, and 9 of the remaining 10 continued illicit drug use. Only one was healthy and off drugs. Over the same period, none of the methadone treatment group suffered major health problems. Thirteen of the original 17 were no longing using illicit drugs. Three more members of the non-methadone group died in the following three years, in a study completed before the era of HIV/AIDS [37].

More recently, treatment for substance abuse appears to lead to a significant reduction in HIV seroconversion among treatment clients. Metzger and colleagues document a sixfold difference in HIV incidence between methadone clients and out-of-treatment IDUs [38]. More recently, analytic models also indicate that methadone maintenance treatment would be highly cost-effective even if its only benefits would be to prevent HIV infection [20, 39, 40]. Using analytic models that examine highly imperfect treatment programmes with high relapse rates, the author of the present paper found that methadone maintenance treatment would

cost between 150,000 and 300,000 United States dollars ($) per averted HIV infection in a high-risk population. These estimates are below lifetime estimated treatment costs [41] and compare favourably with widely accepted interventions to improve and to extend human life [42].

Methadone maintenance treatment and other approaches perform three different and complementary functions to prevent the spread of HIV and other bloodborne diseases.

First and foremost, drug treatment reduces the likelihood and the frequency of injecting drug use. Those modalities are hardly perfect. Substance abuse is a chronic, relapsing condition. Not surprisingly, many clients relapse or fail to comply with recommended treatment. Yet because HIV is difficult to transmit, modest and imperfect programmes can have a large impact on the spread of disease [43]. Though beyond the scope of the present paper, best-practice treatment includes sufficiently high methadone dosage [44] and the provision of accompanying social services to reduce behavioural risks [45, 46].

In addition to reducing drug use, treatment for substance abuse can also include harm reduction elements such as instruction on the proper use of bleach. Because most treatment clients will engage in some level of future drug use, such instruction may have a substantial impact on the spread of infectious disease. These approaches are controversial. Some treatment professionals view instruction on "safer injection" as implicitly condoning injecting drug use [47]. The author of the present paper considers that harm reduction services are essential to protect treatment clients, given the realities of widespread relapse and nonadherence in most treatment settings.

Treatment for substance abuse plays a third role in providing a venue for identifying and delivering medical treatment for infected IDUs. Most substance-abuse treatment facilities in the United States currently provide routine HIV counselling and testing [47]. Many, though by no means all, provide effective linkages to required medical care [48]. Although the impact of such interventions on the spread of disease is not well known, it is plausible that such linkages reduce the severity of the disease and the likelihood of further transmission.

Treatment for substance abuse should be held to high standards as a health-care intervention

Despite the above-mentioned achievements, treatment for substance abuse would be much more effective if its core mission included the provision of high-quality health care. Many treatment providers describe substance abuse as a treatable medical condition. Yet few drug treatment facilities provide high-quality care when judged by best-practice standards of medical care applied in other settings. Many facilities do not provide basic physical examinations, reproductive health services or infectious disease screening for highly prevalent conditions among drug users.

The author of the present paper recently visited a large and reputable United States provider of methadone maintenance. The facility used a high-technology

management system to provide accurate billing and to meet the exacting requirements of government regulators. Yet the same facility dispensed methadone in assembly-line fashion with rudimentary medical and social services to a high-risk population. Quite openly, the facility provided minimal services to uninsured adults for whom it received low rate of reimbursement.

A critical problem is to provide effective treatment for those drug users at greatest risk of spreading and contracting infectious disease. That is most obvious in the area of HIV prevention, where IDUs with dual-diagnosis psychiatric disorders, unmarried young men, those with intense criminal justice involvement, those in poverty and polydrug users are more likely to contract or to transmit HIV [49]. Precisely the same characteristics are correlated with non-adherence and premature exit from treatment for substance abuse [50]. Treatment providers—especially those evaluated on the basis of mean relapse rates and other measures—have strong incentives to select the most cooperative and successful clients and to avoid those most at risk of HIV infection.

More generally, public policies must recognize the unique role of injecting drug use. Although marijuana users and other, relatively low-risk, drug users require appropriate care, resource allocation for substance-abuse treatment must reflect the public health reality that curbing injecting drug use is essential to control blood-borne disease.

Drug treatment facilities are probably much more effective in dealing with HIV than they are in addressing prevention and treatment of other blood-borne diseases. For example, IDUs face high risks of contracting hepatitis B virus (HBV). Many treatment facilities provide rudimentary screening and treatment, despite strong evidence of the efficacy and effectiveness of HBV vaccination among IDUs [51, 52].

Paradoxically, services to address other blood-borne ailments may be especially important in areas of low or declining rates of new HIV infection. Because HIV provides unique motivation to avoid needle-sharing and other high-risk behaviours, additional services may be required to address behavioural risks involving other serious blood-borne ailments that are more infectious, if less prominent, than HIV itself. For that reason, policy makers must be open to the possibility that there is some trade-off between policies that minimize the spread of HIV and those that minimize the spread of other infectious diseases.

SEPs reflect some of the same limitations that are common within substance abuse treatment. Many SEPs have important limitations that hamper both their effectiveness and the political sustainability of this controversial intervention.

For SEPs to curb more infectious agents successfully, programme quality must be improved. Some SEPs have already added drug treatment referrals, case management for HIV prevention and other services that go beyond the distribution of sterile equipment [53]. Such linkages would provide a valuable means of bringing high-risk individuals into treatment. In that connection, analysis by Kim Blankenship of Yale University (United States) indicates that HIV-negative IDUs are ineligible for some of the most important preventive services.

The criminal justice system should be used to prevent and treat disease

Finally, efforts to control infectious disease must find more effective ways to identify and to serve out-of-treatment IDUs who face the greatest risk of disease. As first documented by Hammett and others for the United States, a large proportion of IDUs can be found in prisons and other correctional settings [28]. Seventeen per cent of all AIDS cases in the United States, 36 per cent of tuberculosis cases and 29 per cent of HCV cases occur within the incarcerated population [29]. Many of the diagnosed inmates are IDUs.

Despite the well-documented need for substance-abuse and health-care services, correctional health-care and substance-abuse treatment are frequently unavailable or sub-standard. A 1991 report of the United States General Accounting Office indicates that federal prisons provide appropriate treatment to only 1 per cent of inmates with significant drug problems [54]. Dramatic incidents of medical misconduct highlight the darker possibilities of correctional care [31, 55, 56].

Even larger numbers of IDUs are under supervision of the criminal justice system in the general community—on probation, parole or pretrial release or in other arrangements. More rigorous supervision, referrals and more effective social and medical services are essential to reduce substance use and to improve the well-being of this population [57]. Research by Pollack, Khoshnood and Altice proposes a multifaceted strategy of entitlement security, case management, outreach and prison discharge planning to address those concerns [28, 31]. Recent innovations such as "coerced abstinence" and graduated sanctions may also be effective in deterring relapse among criminal offenders [58].

Conclusion

The AIDS epidemic provides the inevitable context in evaluating public health interventions for IDUs. Like any powerful historical experience, HIV presents powerful lessons that should be heeded in confronting other blood-borne epidemics.

One clear lesson is that disease prevention requires effective public management and carefully applied epidemiology. The late and inadequate United States response to HIV reflected public indifference to a problem afflicting IDUs and other stigmatized groups [59]. Yet that policy failure also reflected the failure of disease surveillance systems to detect HIV prevalence among IDUs quickly, the failure of policy makers to fund and implement substance abuse treatment adequately and the failure of health-care delivery and criminal justice systems to implement best-practice HIV prevention among men and women who face the greatest behavioural risks [13].

Many of the policy failures could have been avoided through improved public management and through the decisive political leadership required to enact controversial policies. One task of such leadership is to teach a skeptical electorate

that public health measures for IDUs are sound public investments. When methadone maintenance treatment or SEPs are evaluated by widely accepted standards, they compare quite favourably with other public health interventions [42]. It would be naive to think that academic studies are sufficient to mobilize political support for unpopular policies. But it would also be wrong to discount the importance of judicious analysis in providing policy makers with the data—at times the political cover—to take unpopular steps. Scientific consensus supporting the effectiveness of SEPs has been cited by a former Mayor of New York City, a former United States Secretary of Health and Human Services and other policy makers. More recently, politicians and commentators across the political spectrum cite evidence that supply-side cocaine enforcement is less cost-effective than interventions for prevention and treatment interventions [36].

The HIV epidemic also serves as a reminder that IDUs are members of the broader community, and are not some incorrigible group beyond help or justified concern. Although IDUs engage in illegal, and sometimes destructive, behaviours, HIV prevention efforts indicate that IDUs possess unexpected capacities to help themselves and others in avoiding deadly disease. Those capacities are an important asset for future interventions [12, 60].

One of the most important lessons is that effective interventions must include elements of both prevention and treatment to be fully effective. To attract clients and to reduce long-term risk, harm-reduction interventions must include conduits to substance-abuse treatment and medical and social services. To provide lasting protection for a client population prone to both relapse and non-adherence to treatment, traditional treatment for substance abuse must include preventive measures such as instruction in safer injecting. For diseases that can be prevented by vaccination, such as HBV, SEPs and substance-abuse treatment also provide opportunities for more permanent prevention. The prevention component of treatment will be equally important for ailments that involve the prospect of recovery and subsequent infection.

Like any powerful experience, HIV teaches some lessons that may prove misleading in addressing other epidemics. Like generals preparing to fight the last war, prevention specialists who uncritically generalize from HIV may therefore be cruelly disappointed. HIV is a uniquely feared infectious agent, is currently impervious to cure or vaccination, and is rather difficult to transmit. Infectious diseases that do not share those characteristics are likely to require different interventions for treatment and prevention.

Opponents of SEPs have long argued that harm reduction measures to make substance use safer are an inadequate response to the individual and social harms associated with injecting drug use. Those critics are wrong about HIV, but they have a more persuasive case concerning more easily transmitted agents that are more difficult to control through imperfect behavioural interventions.

Put differently, the spread of HCV and other infectious agents provides a painful reminder of both the necessity and the limitations of harm reduction. Measures such as SEPs reduce the risk of disease and provide critical outreach to engage the hidden population of IDUs. Those modest interventions are not enough

to protect men and women in an environment of persistent and frequent needle-sharing and other behavioural risks. Future harm-reduction measures may therefore require greater and more explicit use reduction than is currently implemented in SEPs. Measures such as SEPs must be further expanded to address other concerns, including overdoses and the management of chronic disease in an aging population of IDUs.

Clinicians and policy makers must also realize that the unique nature of the HIV epidemic created new opportunities that are unlikely to be replicated in the fight against other diseases. Among those at personal risk, fear is a powerful motivator. HIV therefore produced remarkable behaviour change among many IDUs who are probably less fearful of other blood-borne diseases [61]. A similar pattern has been observed among gay and bisexual men, some of whom have resumed high-risk behaviours in response to declining local HIV incidence and the overall development of more effective HIV therapies [10, 62].

Within the broader society, HIV prevention is a major and highly visible social issue, one that links the interests of IDUs with those of other citizens at risk of disease. In many countries, the HIV epidemic framed prevention and treatment measures for IDUs as part of a broader national effort to contain a new and frightening infectious disease. HIV greatly increased the amount of public attention and public resources devoted to substance-abuse policy. More obscure and less lethal blood-borne ailments are unlikely to produce such an expensive and focused policy response.

The spread of blood-borne disease is a serious public health threat among IDUs. The threat will remain, even in the happy event that HIV is ever eradicated within that population. Even if the right lessons are learned from HIV, not all problems associated with the spread of blood-borne disease will be solved. But it will be possible to confront them more effectively, and thereby reduce the avoidable suffering being experienced by those who continue to inject illicit drugs.

References

1. R. S. Garfein and others, "Viral infections in short-term injection drug users: the prevalence of the hepatitis C, hepatitis B, human immunodeficiency, and human T-lymphotropic viruses," *American Journal of Public Health,* vol. 86, No. 5 (1996), pp. 655-661.
2. R. S. Garfein and others, "Prevalence and incidence of hepatitis C virus infection among young injection drug users", *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, vol. 18, Supplement 1 (1998), pp. S11-S19.
3. C. Contoreggi, V. E. Rexroad and W. R. Lange, "Current management of infectious complications in the injecting drug user", *Journal of Substance Abuse Treatment,* vol. 15, No. 2 (1998), pp. 95-106.
4. N. Salomon and others, "Prevalence and risk factors for positive tuberculin skin tests among active drug users at a syringe exchange program", *International Journal of Tuberculosis and Lung Disease,* vol. 4, No. 1 (2000), pp. 47-54.

5. E. Kaplan, and D. Soloschatz, "How many drug injectors are there in New Haven? Answers from AIDS data", *Mathematical and Computer Modelling,* vol. 17, No. 2 (1993), pp. 109-115.
6. E. Stokey and R. Zeckhauser, *A Primer for Policy Analysis* (New York, Norton, 1978).
7. J. Hodges, "Six (or so) things you can do with a bad model", *Operations Research,* vol. 39, No. 3 (1991), pp. 355-365.
8. E. H. Kaplan and H. A. Pollack, "Allocating HIV prevention resources", *Socio-Economic Planning Sciences,* vol. 32, No. 4 (1998), pp. 257-263.
9. E. Kaplan, "Economic evaluation and HIV prevention community planning: A policy analyst's perspective", in *Handbook of Economic Evaluation of HIV Prevention Programs,* D. Holtgrave, ed. (New York, Plenum, 1998).
10. Institute of Medicine, *No Time to Lose: Getting More From HIV Prevention* (Washington, D.C., National Academy Press, 2000).
11. D. Altman, R. Greene and H. Sapolsky, *Health planning and regulation: The decision-making process* (Ann Arbor, AUPHA Press, 1981), p. 81.
12. D. Des Jarlais, S. Friedman and T. Ward, "Harm reduction: a public health response to the AIDS epidemic among injecting drug users", *Annual Review of Public Health,* vol. 14, 1993, pp. 413-450.
13. D. Des Jarlais and others, "Maintaining low HIV prevalence in populations of injecting drug users", *Journal of the American Medical Association,* vol. 274, No. 15 (1995), pp. 1226-1231.
14. Institute of Medicine, *Preventing HIV Transmission: the Role of Sterile Needles and Bleach* (Washington, D.C., National Academy Press, 1995).
15. J. K. Watters and others, "Syringe and needle exchange as HIV/AIDS prevention for injection drug users", *Journal of the American Medical Association,* vol. 271, No. 2 (1994), pp. 115-120.
16. P. Lurie and others, *The Public Health Impact of Needle Exchange Programs in the United States and Abroad* (San Francisco, Institute for Health Policy Studies, University of California, 1993).
17. E. Kaplan, "A circulation theory of needle exchange", *Operations Research,* vol. 43, No. 4 (1995), pp. 558-569.
18. A. Goldstein, "Clinton refuses needle exchange funding", *Washington Post,* 28 April 1998.
19. United States of America, Department of Health and Human Services, "Research shows needle exchange programs reduce HIV infections without increasing drug use" (Washington, D.C., 20 April 1998).
20. H. A. Pollack, "Can we protect drug users from hepatitis C?", *Journal of Policy Analysis and Management,* vol. 20, No. 2 (2001), pp. 358-364.
21. A. Moss and J. Hahn, "Needle exchange: no help for hepatitis?", *American Journal of Epidemiology,* vol. 149, No. 3 (1999), pp. 214-216.
22. H. Hagan and others, "Syringe exchange and risk of infection from hepatitis B and C viruses", *American Journal of Epidemiology,* vol. 149, No. 3 (1999), pp. 203-213.
23. R. Coutinho, "HIV and hepatitis C among injecting drug users: success in preventing HIV has not been mirrored for hepatitis C", *British Medical Journal,* vol. 317, 1998, pp. 424-425.

24. N. Crofts and others, "Blood-borne virus infections among Australian injecting drug users: implications for spread of HIV", *European Journal of Epidemiology,* vol. 10, No. 6 (1994), pp. 687-694.

25. N. Crofts and others, "Methadone maintenance and hepatitis C virus infection among injecting drug users", *Addiction*, vol. 92, No. 8 (1997), pp. 999-1005.

26. E. Kaplan, E. Kedem and S. Pollack, "HIV incidence in Ethiopian immigrants to Israel", *Journal of AIDS,* vol. 17, No. 5 (1998), pp. 465-469.

27. R. Janssen and others, "New testing strategy to detect early HIV-1 infection for use in incidence estimates and for clinical and prevention purposes", *Journal of the American Medical Association,* vol. 280, No. 1 (1998), pp. 42-48.

28. T. M. Hammett, J. L. Gaiter and C. Crawford, "Researching seriously at-risk populations: health interventions in criminal justice settings", *Health Education and Behavior*, vol. 25, No. 1 (1998), pp. 99-120.

29. T. Hammett, W. Rhodes and P. Harmon, "HIV/AIDS and other infectious diseases among correctional inmates: a public health problem and opportunity" (abstract 571), in *National HIV Prevention Conference 1999* (Atlanta, Centers for Disease Control and Prevention, 1999).

30. A. Thompson and others, "Evaluation of an innovative program to address the health and social service needs of drug-using women with or at risk for HIV infection", *Journal of Community Health,* vol. 23, No. 6 (1998), pp. 419-440.

31. H. A. Pollack, K. Khoshnood and F. Altice, "Health-care delivery strategies for criminal offenders", *Journal of Health Care Finance,* vol. 26, No. 1 (1999), pp. 63-78.

32. D. R. Gerstein and H. J. Harwood, *Treating Drug Problems* (Washington, D.C., National Academy Press, 1990).

33. M. D. Anglin and D. G. Fisher, "Survival analysis in drug program evaluation. Part II—Partitioning treatment effects", *The International Journal of the Addictions,* vol. 22, No. 4 (1987), pp. 377-387.

34. J. C. Ball and A. Ross, *The Effectiveness of Methadone Maintenance Treatment: Patients, Programs, Services, and Outcome* (New York, Springer-Verlag, 1991).

35. P. Barnett, "The cost-effectiveness of methadone maintenance as a health care intervention", *Addiction,* vol. 94, No. 4 (1999), pp. 479-488.

36. J. Caulkins, "Mandatory minimum drug sentences: throwing away the key or the taxpayers' money?" (Santa Monica, California, RAND Drug Policy Research Center, 1997).

37. L. Gunne and L. Gronbladh, "The Swedish methadone maintenance program: a controlled study", *Drug Alcohol Depend,* vol. 7, No. 3 (1981), pp. 249-256.

38. D. Metzger and others, "Human immunodeficiency virus seroconversion among intravenous drug users in- and out-of-treatment: an 18-month prospective follow-up", *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology,* vol. 6, No. 9 (1993), pp. 1049-1056.

39. G. S. Zaric, M. L. Brandeau and P. G. Barnett, "Methadone maintenance and HIV prevention: a cost-effectiveness analysis", *American Journal of Public Health,* vol. 90, No. 7 (2000), pp. 1100-1111.

40. G. S. Zaric, P. G. Barnett and M. L. Brandeau, "HIV transmission and the cost-effectiveness of methadone maintenance", *Management Science,* forthcoming.

41. D. Holtgrave and S. Pinkerton, "Updates of cost of illness and quality-of-life estimates for use in economic evaluations of HIV prevention programs", *Journal of AIDS,* vol. 16, No. 1 (1997), pp. 54-62.

42. T. O. Tengs and others, "Five-hundred life-saving interventions and their cost-effectiveness", *Risk Analysis,* vol. 15, 1995, pp. 369-390.

43. H. A. Pollack, "Methadone maintenance as HIV prevention: cost-effectiveness analysis", in *Quantitative Evaluation of HIV Prevention Programs,* E. H. Kaplan and R. Brookmeyer, eds. (New Haven, Connecticut, Yale University Press, forthcoming).

44. T. D'Aunno and T. Vaughn, "Variations in methadone treatment practices: results from a national study", *Journal of the American Medical Association,* vol. 267, No. 2 (1992), pp. 253-258.

45. S. Avants and others, "Day treatment versus enhanced standard methadone services for opioid-dependent patients: a comparison of clinical efficacy and cost", *American Journal of Psychiatry,* vol. 156, No. 1 (1999), pp. 27-33.

46. E. C. Strain and M. L. Stitzer, *Methadone Treatment for Opioid Dependence* (Baltimore, Johns Hopkins University Press, 1999).

47. T. D'Aunno, T. Vaughn and P. McElroy, "An institutional analysis of HIV prevention efforts by the nation's outpatient drug abuse treatment units", *Journal of Health and Social Behavior,* vol. 40, No. 2 (1999), pp. 175-192.

48. P. Friedmann and others, "Medical and psychosocial services in drug abuse treatment: do stronger linkages promote client utilization?", *Health Services Research,* vol. 35, No. 2 (2000), pp. 443-465.

49. D. Metzger and others, "Risk factors for needle-sharing among methadone-treated patients", *American Journal of Psychiatry,* vol. 148, No. 5 (1991), pp. 636-640.

50. S. Magura, P. C. Nwakeze and S. Demski, "Pre- and in-treatment predictors of retention in methadone treatment using survival analysis", *Addiction,* vol. 93, No. 1 (1998), pp. 51-60.

51. L. Borg and others, "Methadone-maintained former heroin addicts, including those who are anti-HIV-1 seropositive, comply with and respond to hepatitis B vaccination", *Addiction,* vol. 94, No. 4 (1999), pp. 489-493.

52. K. Seal and others, "Risk of hepatitis B infection among young injection drug users in San Francisco: opportunities for intervention", *Western Journal of Medicine,* vol. 172, No. 1 (2000), pp. 16-20.

53. R. Heime, "Can syringe exchange serve as a conduit to substance abuse treatment?", *Journal of Substance Abuse Treatment,* vol. 15, No. 3 (1998), pp. 183-191.

54. United States of America, General Accounting Office, "Drug treatment: despite new strategy, few federal inmates receive treatment", report No. GAO/HRD92-116 (Washington, D.C., Government Printing Office, 1991).

55. M. Puisis, "Clinical practice in correctional medicine" (Saint Louis, Missouri, Mosby, 1998).

56. A. Skolnick, "Critics denounce staffing jails and prisons with physicians convicted of misconduct", *Journal of the American Medical Association,* vol. 280, No. 16 (1998), pp. 1391-1392.

57. M. A. Kleiman, *Against Excess: Drug Policy for Results* (New York, Basic Books, 1993).

58. M. A. Kleiman, "Drugs and drug policy: the case for a slow fix", *Issues in Science and Technology,* vol. XV, No. 1 (1998).

59. C. Cohen, *The Boundaries of Blackness: AIDS and the Breakdown of Black Politics* (Chicago, University of Chicago Press, 1999).

60. R. Broadhead and others, "Harnessing peer networks as an instrument for AIDS prevention: results from a peer-driven intervention", *Public Health Reports,* vol. 113, supplement 1 (1998), pp. 42-57.

61. R. Battjes, R. Pickens and L. Brown, "HIV infection and AIDS risk behaviors among injection drug users entering methadone treatment: an update", *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology,* vol. 10, 1995, pp. 90-96.

62. G. Rotello, *Sexual Ecology: AIDS and the Destiny of Gay Men* (New York, Plume, 1998).

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

Age-specific multi-state drug initiation models: insights from considering heterogeneity*

C. ALMEDER
Department of Operations Research and System Theory, Vienna University of Technology, Vienna, Austria

J. P. CAULKINS
Drug Policy Research Center, RAND Corporation, Santa Monica, California, United States of America

G. FEICHTINGER AND G. TRAGLER
Department of Operations Research and System Theory, Vienna University of Technology, Vienna, Austria

ABSTRACT

The present paper introduces several models concerning drug initiation and drug epidemics that extend traditional single and multi-state dynamic models by explicitly considering the age distribution of users. This age-specificity yields more realistic models, because both human behaviour and the influence of personal relationships can depend on age and the age difference. Furthermore, prevention programmes—especially school-based programmes—can be targeted to reach certain age groups. The models allow the dynamics of drug epidemics to be reproduced in greater detail. The models can be used either for studying the features of drug initiation or for discovering how best to allocate resources to prevention programmes for different age groups.

Introduction

Models of drug initiation and drug use are often based on the same principles as epidemiological models. At first sight, it is not clear why drug use and infectious diseases have much in common. Initiation into consumption of an illicit drug is usually a deliberate decision of an individual, whereas the act of infection by a pathogenic agent (for example, the common cold virus) usually takes place without the infected individual having any awareness of the process. In contrast to a drug epidemic, most individuals infected by a certain disease undergo the same run of infection, that is, almost all individuals either recover or become immune

*Research for the present paper was financed in part by the Austrian Science Foundation.

or remain infected. The process of recovery from drug consumption may differ for different users. It may be a healing process, if the user was addicted, or may result from the user's free choice to quit consumption or from the application of drug enforcement policies [1].

Nevertheless, drug use is clearly contagious in the sense that use by some individuals affects the probability that others will use it through multiple mechanisms. In a very literal sense, most users are introduced to a drug by a friend or relative; the more drug users there are, the more likely an individual is to be offered the drug [2]. At a market level, the larger the market, the more diluted the enforcement risk, and the safer it is to try drugs [3]. At a reputational level, experiences of others can be instrumental in shaping perceptions of the riskiness of drugs, and those perceptions in turn influence initiation [4]. Indeed, the very fact that the mechanism of transmission does not involve physical contact or interaction means that the dynamics of contagion can be more complex and more interesting.

Looking at drug use and the process of initiation in more detail, it is clear that the decision of a non-user to start consumption depends strongly not only on the immediate, personal social environment of the person concerned, but also on the overall reputation of a drug in society, for example, as portrayed in movies or news media. Therefore, a person might want to use drugs even though none of that person's associates encourages the desire. Conversely, a person may fear drugs even if no one that he or she knows has suffered harm from them.

The foregoing rough explanation already shows the complexity of drug initiation and the large number of parameters influencing the process. The complexity grows if the age dependency of those factors is taken into consideration. The disadvantages of complex models stem from the lack of data for parameter identification and the difficulties in analysing them. Hence, in the present paper, as age dependency is introduced into models of drug use, an effort will be made to keep other aspects of the models relatively simple.

More specifically, the additional insight considered and quantified in the present paper is that the influence of a drug user on a non-user vulnerable to initiation can depend on the ages of both persons. A 16-year-old might look up to and seek to emulate an 18-year-old, but rebel and try to do exactly the opposite of what his or her parent's generation is perceived to be doing. So drug use by an 18-year-old might encourage a 16-year-old to initiate use, even though drug use by a 38-year-old might discourage it. Or perhaps for some drugs and some situations, use by a 38-year-old is a stronger endorsement than is use by an 18-year-old. Whichever scenario is considered likely, there is no way to incorporate such effects into existing models. There are age-specific epidemiological models in the literature [5–6], but, as mentioned above, the process of contagion is different for drug initiation. Therefore, those models are not capable of describing the dynamics of a drug epidemic. The present paper describes a new class of models that can consider such effects in an explicit quantitative framework and gives some initial results. Furthermore, control models of prevention, which are based on the descriptive age-specific drug initiation models, are introduced.

Basic aspects of drug initiation models

First of all, a very general description of the basic features of drug initiation models is given. Those features are the foundation for the further development of initiation models towards age-structured models, which are presented below (see "Descriptive age-specific models of initiation"). Some heterogeneity is already introduced by dividing the population into multiple groups (see "Multi-state models" below), but that is qualitatively different from the models, introduced later, with an infinite number or continuous distribution of states.

Homogeneous modelling of the population—single-state model

The simplest model of drug use in a population distinguishes only between persons who do not consume drugs and those who do. It is common in such models to ignore variation in the overall population size. That means that the number of deaths is equal to the number of births. In that case, the only things to consider are the flows from the non-user group to the user group and back (see figure I). The flow from the non-user to the user group—the initiation rate—is influenced by several factors, including the following:

(a) A basic initiation rate representing the probability that a non-user starts drug consumption without any influence from others. The basic initiation rate is drug-specific, but also culture-specific. Therefore, different initiation rates must be considered not only for different drugs, but also for different cultural contexts. Apart from that, the basic initiation rate may also depend on time (in the United States of America, for example, there were higher basic initiation rates for cocaine during the late 1960s than during the early 1950s);

(b) The reputation of the drug in the society. Reputation is determined by the number of drug consumers, but it is not necessarily the case that more drug users always mean a more favourable reputation. For some drugs, there could be a threshold. Beneath that threshold, the reputation increases with the number of users, but above the threshold, it decreases because a very high number of drug users could underscore the problems caused by drug consumption;

(c) The effect of prevention programmes. Prevention programmes can reduce the initiation rate, but only down to a certain level. Unfortunately, no programme can reach all people, and not all people reached change their behaviour.

The flow from the user group to the non-user group also involves the following different factors:

(a) The flow implied by the assumption of zero population growth (if all newborns are assumed to be in the non-user group, then there must be a flow from the user to the non-user group representing the death of users and their replacement by newborns who are not users);

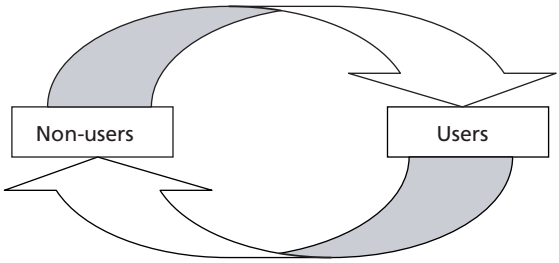
(b) A basic rate at which drug users stop their consumption;

(c) The effect of treatment. With more treatment the flow from users to non-users is increased, but only up to a certain threshold (similar to the effects of prevention programmes on the initiation rate).

Note that if the user group represents all people who have consumed drugs at least once in their lifetime, then the flow from the user to the non-user group consists only of the part due to the zero population growth.

When considering models with a constant population size, it is necessary to analyse only one of the groups; all individuals who are not in that group must be, by definition, in the other. The dynamics of the model therefore depends only on one state, and the analysis becomes accordingly simpler.

Figure I. Simple user—non-user model



Multi-state models

One type of a more detailed model distinguishes not only between drug users and non-users, but also between different kinds of users and possibly also between different kinds of non-users. Such a multi-state model is explained by Everingham and Rydell [7], who divide users into groups of light and heavy users. The light drug users consume drugs at most once per week; the heavy users consume several times per week or more often.

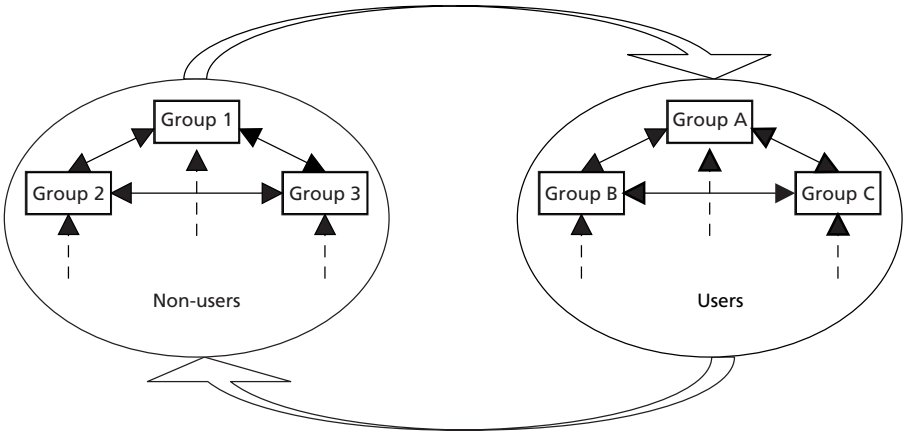
Even that simple binary distinction generated many insights, but there are other ways of splitting the population in multiple groups. A principle constraint on such elaborations is data availability. For example, if there is no distinction in the data between the amounts used and the frequency with which people use drugs, then it makes no sense to introduce different levels of drug consumption into the models.

Although the population is split into several groups, they can all be combined again into two sets of groups, namely those groups representing the non-user population and those representing the drug consumers (see figure II). Some examples of groups within the non-user population are: general non-users; non-users who have been exposed to a prevention programme; and former users. Examples of groups within the user population are: light users; heavy users; and users under treatment.

In the multi-state models, there are not only transitions between non-users and users, but also within those groups, for example, from light users to heavy

users. Very complex models can be generated with many different states, but, as mentioned above, it is only useful to develop a model that can be supported by the available data. Highly differentiated data acquisition allows modelling of drug dynamics in greater detail.

Figure II. Multi-stage model with subgroups of users and non-users



Descriptive age-specific models of initiation

The central contribution of the present paper is to introduce another form of heterogeneity to the models, specifically to take into consideration the age of the individuals. Such an extension yields much more realistic models, because it is certainly true that behaviour can depend on age. For example, people between 13 and 25 years of age are much more likely to start drug consumption than are people over 40. It is also true that not all people have the same influence on a person of a certain age. That influence can depend on their age and the age difference. Furthermore, prevention programmes—especially school-based programmes—can be targeted to reach certain age groups. Therefore, such an extension is useful, although the increased complexity makes it much more difficult to analyse such models and gain practical results [8].

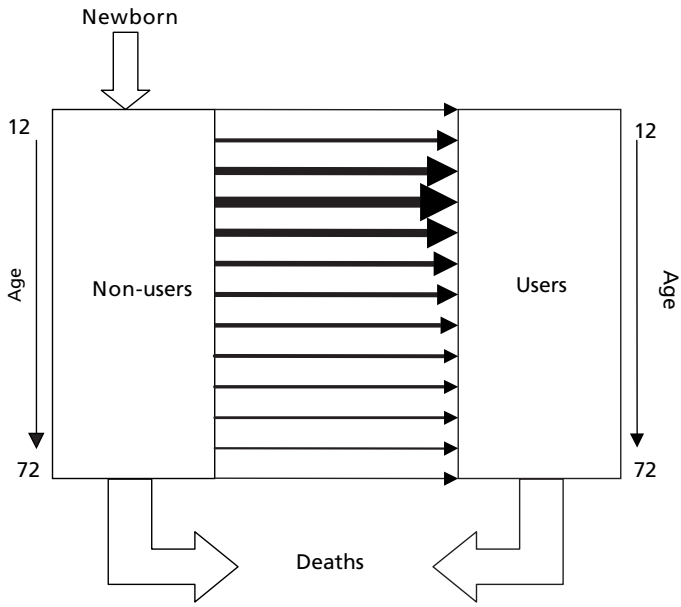
A simple way to introduce age is to split the population into different age groups. That leads to so-called compartment models, such as those described above (see "Multi-state models"), but with a large number of population groups [9]. A more general method is to include age as a second parameter in addition to time. A continuous age distribution of the population can then be fully considered. The analysis of the model becomes in some way easier, because there are fewer groups to consider, but in some sense also more complicated, because the method leads to a system of partial differential equations—a further development of the so-called McKendrick equation [10].

Age-specific single-state models

As a first step, the simplest model with 2 groups (that is, non-users and users) is extended to consider age-specific reputational feedback on initiation. The details of the underlying model change a little. It is assumed that there is a fixed birth cohort size, which moves only into the non-user group at age 12 (that is, any use before the age of 12 is ignored). There are no deaths during the subsequent lifetime of 60 years, so that all individuals live from age 12 to age 72. Then they are removed from the model, either by death or because they are presumed no longer to affect drug use. In addition, it is assumed that there is no recovery from drug use, so that the backflow from users to non-users can be neglected (see figure III). Since people obviously do cease drug use, the practical implication of the simplification is that the pool of users is more properly understood to be the pool of those who ever used drugs, not the pool of current or past-year users.

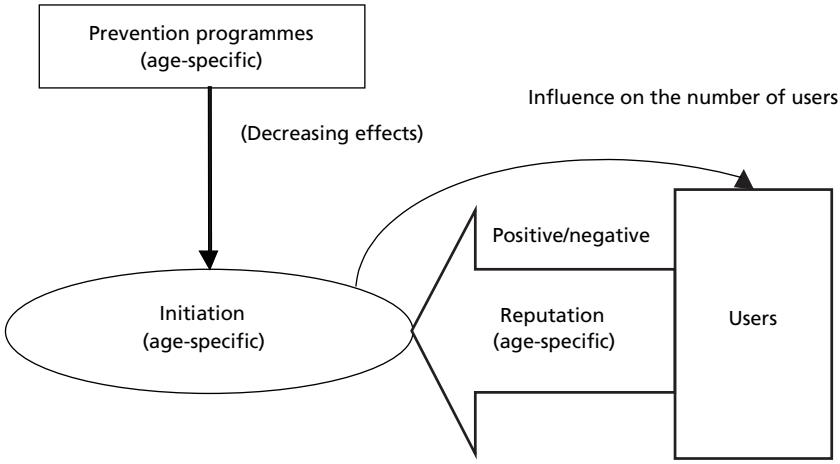
The initiation rate involves three factors (see "Homogeneous modelling of the population—single-state model" above), namely the basic initiation rate, the reputation of the drug and the effect of prevention programmes. Each factor depends on age. Prevention programmes always suppress initiation, but the reputation effect can influence initiation in either direction (see figure IV).

Figure III. Schematic representation of a simple age-specific model



Note: Thicker arrows indicate higher initiation rates.

Figure IV. Influences on the initiation rate



In the simpler model without age differentiation, the reputation was construed as an overall image held by the society as a whole, but a more differentiated view is now adopted because the reputation is age-specific. In other words, at any given point in the drug epidemic, people of different ages can have different perceptions of the drug. Furthermore, the influence of a given user on someone else's perceptions of the drug can depend on age and the age difference. So the main factor for the reputation is how much influence a user of age b has on a non-user of age a, and whether the influence is positive or negative. Such a relationship can be a socially- or culturally-specific factor. The authors believe that, at least for the typical drug in North America or western Europe, such influence may be described as follows:

(a) Young people are more sensitive to peer pressure, so that the older a non-user is, the weaker the influence of any other user on that individual;

(b) The influence depends primarily on the age difference between the user and non-user;

(c) Especially for young non-users (who have the highest underlying proclivity to initiate drug use):

(i) Persons who are of the same age or a little older set examples, and their influence is therefore very high;

(ii) Persons who belong to their parents' generation may evoke a contrarian response, or at least have only a small impact, and therefore a very low and possibly even a negative influence.

The reputation of the drug that is effective on a non-user of age a is the compound influence of the users over all ages.

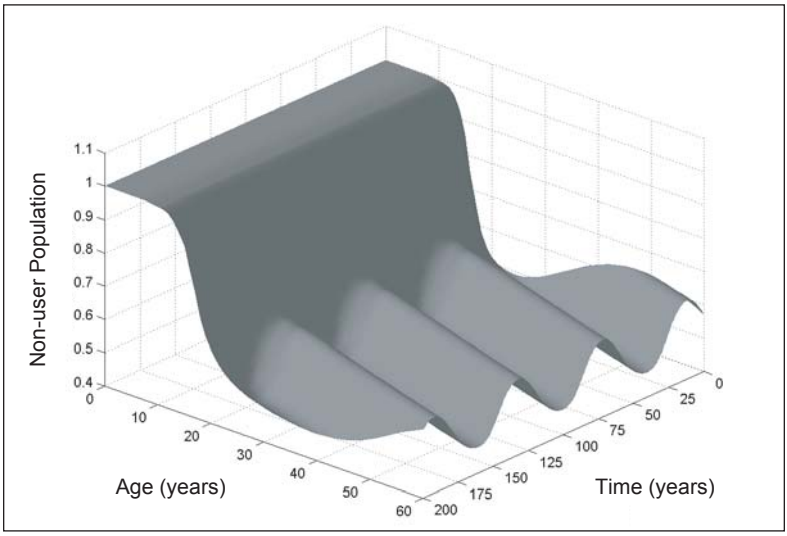
The feedback from considering the influence of users on the initiation rate of non-users can generate very interesting behaviour in the model. The drug users never vanish completely, because neither a low reputation, nor an expensive

prevention programme can reduce initiation to zero. In most cases, when no prevention programme is applied, the number of users becomes stabilized and reaches an endemic equilibrium. On the other hand, it is also possible to end up with an oscillation of repeated waves of drug use, if the influence of the parents' generation is negative.

The intuitive explanation for the modelled behaviour is the following. If there are few older drug users, then their negative influence on the younger potential users is very small. Therefore, initiation of new drug users at younger ages is high (initiation by older individuals is low and relatively insensitive to use by others). Some years later, those former young drug users grow older and cause a strong negative reputation of the drug for the new young non-users. That keeps the initiation rate very low and results in a large non-user group in younger and eventually middle ages. The older users eventually disappear, which reduces their negative influence on initiation, and the cycle can begin again.

Figure V shows the cycles graphically. In particular, it plots, as a function of time, the proportion of an age cohort that has never used drugs. Troughs in the graph represent peaks in use because everyone is either a user or a non-user. People in the youngest age groups are primarily non-users at all times, as is indicated by the plateau in the graph for those ages. Initiation occurs primarily in the adolescent years (indicated by the cliff at those ages), but the ultimate lifetime prevalence of drug use varies from birth cohort to birth cohort (indicated by the series of ridges). The time differences between peaks of drug use are about one to two generations, depending on the influence of users on non-users. The amplitude is about 10 to 20 per cent of a birth cohort size.

Figure V. Oscillation in the non-user population due to the negative influence of the parents' generation



Note: The results are based on data on marijuana initiation in the United States. The user group represents those people who have consumed marijuana at least once in their lifetime.

As mentioned before, the reputation factor depends strongly on the social structure, that is, it may have completely different forms for drug epidemics in other countries or times. The example shows, however, that a model with just one state but a heterogeneous age structure can produce cycles of drug use of the type observed historically [11]. Furthermore, such cycling is not the only possible outcome. It is therefore possible to explore conditions under which cycling occurs and contrast them with conditions under which drug use approaches a constant steady state. Still more complex behaviour can be obtained by further refining the state space, as discussed below.

Age-specific multi-state models

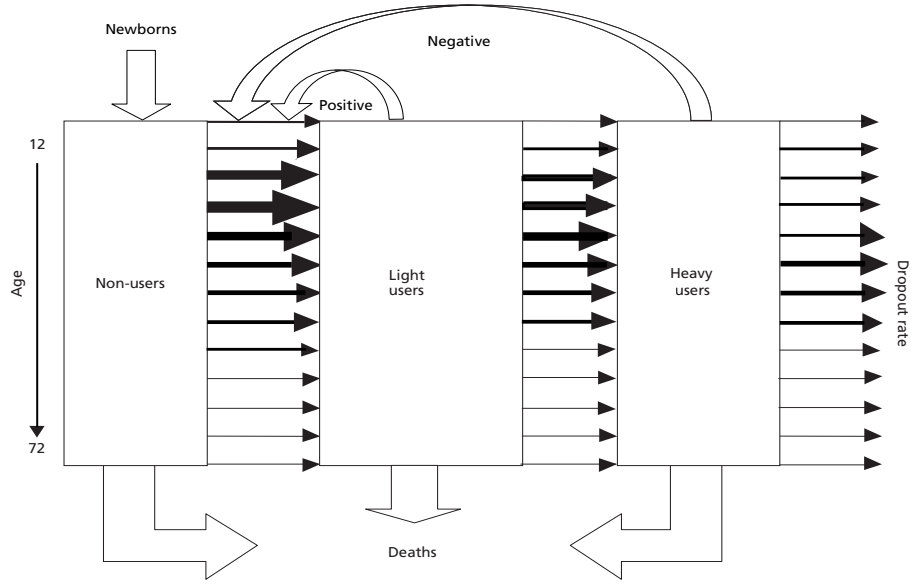
One way to refine the model described above is by splitting the user and non-user groups into subgroups as shown above (see "Multi-state models"). Following the lead of Everingham and Rydell [7] and Behrens and others [12], a distinction is made between light and heavy users. That yields three population groups consisting of non-users, light users and heavy users [13]. A key motivation for the distinction is the assumption that light users, who may be using the drug for recreational reasons, can create a positive impression of the benefits of drug use, whereas the heavy or addicted users who manifest the adverse effects of drug use create a negative impression. Hence, the underlying assumption of the model is that the more light users there are, the better the reputation of the drug is, but the more heavy users there are, the worse the drug's reputation and, hence, the lower the initiation rate (see figure VI).

As the model includes age, the reputation is age-specific and the main factor is again the impact of light and heavy users of age a on the initiation rate of a non-user aged b. A big influence is assumed if the age difference is small, and a small impact if the age difference is large, which reflects the fact that people have more social contacts within their own age group than with other age groups.

Similar to the model discussed above (see "Age-specific single-state models"), the L-H-model shows different types of behaviour depending on the magnitude of the negative influence of heavy users on reputation. If the negative influence is very small, then drug use converges to a single, high equilibrium level of drug use. If it is very large, then initiation is uniformly low and use converges to a single, low equilibrium level. However, if the influence of heavy users is in a moderate range, the model produces cycles of greater and lesser drug use [13]. Intuitively, the reason is about the same as for the model presented above (see "Age-specific single-state models"). A high number of heavy users reduces the initiation rate for younger non-users, but as they grow older and disappear, their negative impact wanes and a new wave of drug consumers is created, which results later on in a high number of heavy users again.

As mentioned above (see "Multi-state models"), that is only one way to extend the model. Other groups can be included as long as data are available for estimating the age-specific transition rates between those groups.

Figure VI. Schematic diagram of an age-specific L-H-model



Note: Thicker arrows indicate higher initiation and transition rates.

Optimization of age-specific models of prevention

The models discussed above were descriptive. Their main contribution was to show how age-specific reputational effects can generate cycles of greater and lesser drug use. From a policy perspective, the aim is not only to describe drug, but also to control them. Since resources are always constrained, a common and fruitful question is to ask how drug control interventions should be managed in order to achieve some reduction in use at the least possible cost.

There is an emerging literature on resource allocation questions, but to date much has focused on trading off spending on different types of interventions (for example, treatment versus enforcement [14-15]), or trading off interventions at different points in a drug control epidemic [16-17], or trading off the benefits of attacking different local markets [18-19]. For school-based prevention programmes, another interesting variant of the question pertains to the appropriate age at which to intervene.

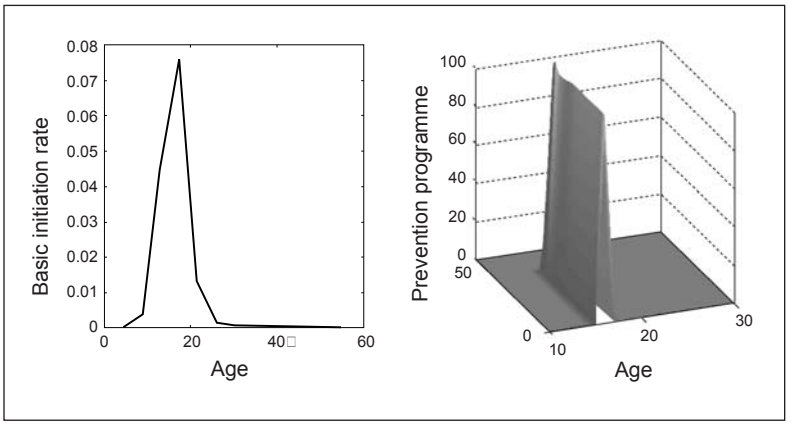
By and large enforcement and treatment interventions cannot be age-specific. Different sanctions can be imposed on juvenile and adult offenders, and treatment policies that are contingent on the number of past treatment episodes have some power of age discrimination, but for school-based drug prevention the ability to target particular ages is much greater. Except for the extent to which students skip or fail grades, there is almost a one-to-one mapping between grade level and age. It can therefore be asked at which grade (age) should most school-based intervention efforts be targeted.

The question is in fact complex and multifaceted. In theory, the preference would be to factor in differential school attendance by age (due to dropping out and truancy) and how intervention varies in effectiveness with age. Consideration is here given to just a subset of those issues that cannot be analysed directly without an age-heterogeneous model. How should the fact that baseline initiation rates vary over time, and that past initiates can promote further initiation, in particular of people from the same birth cohort, affect the ages of intervention?

To address that question, the previously introduced models are converted into an optimal dynamic control formulation [13]. A variety of objectives can be considered, for example, the total costs of drug use, including the social costs caused by the drug consumers and the costs of running the prevention programmes themselves. Initially, it was assumed that a prevention programme only affects the initiation rate while the programme is being run. When the programme is stopped, the initiation rate immediately returns to the rate that would have pertained in the absence of any prevention programme. For that simple case, in which the prevention programmes have no memory effects, the results are predictable. In general, expenditures should be directed to the age classes with the highest initiation rates (see figure VII). However, the spending profile does not match the initiation profile exactly. Because of the contagious character of drug use, prevention interventions should be shifted somewhat towards younger ages.

One way to include prevention programmes with a memory effect is to split the non-users group into two parts, a group of people who have already taken part in a prevention programme, and another group that has not. So the implementation of a prevention programme results in a flow from one of those groups into another. A backwards flow could be included to represent forgetting or a decay in

Figure VII. Basic initiation rates and prevention programmes



Note: The graph on the left shows the age-specific basic initiation rate, which was used for the calculations. The graph on the right shows a typical result for an age-specific prevention programme without a memory effect. The small age range of the prevention programme emerges from the relatively low social costs of drug users. Assuming higher social costs provokes a wider age range.

the effects of the programme. People who have taken part in a prevention programme would presumably be modelled as having a reduced drug initiation rate compared to those who have not taken part. Such a model can be compared to some extent with vaccination models, where the vaccinated part of the population is in one group. After some time, the vaccination effects fade away and people are again susceptible to infection.

The model falls into the class discussed above (see "Age-specific multi-state models"), and for such a prevention programme an optimization can be performed with regard to some objective. Hence, the most important age classes for applying a prevention programme can be calculated, so that the best effects can be obtained (the analysis of the optimal control models is currently under way).

Conclusion

Introducing age-specific aspects into drug initiation models allows the development of more complex models that give detailed insight into the underlying processes of drug epidemics and their control. The simple single-state model discussed above (see "Age-specific single-state models") already shows a big advantage of age-structured models: the complex dynamics of a drug epidemic can be simulated using a model based on simple, manageable assumptions, such as distinguishing between just two groups of people. The behaviour of the solution (cycles or constant equilibrium) depends on the type of age-specific feedback. To gain similar results with traditional multi-state models that do not differentiate by age, a larger number of groups would be necessary.

In particular, the age-specific concept allows the incorporation of an age-specific reputation effect (feedback from the number of users on the initiation rate), which depends on how much influence a user of age b has on a non-user of age a, and whether that influence is positive or negative.

Another advantage of age-specificity is the ability to investigate age-structured strategies for prevention and to optimize them. Although only programmes without a memory effect are addressed above (see "Optimization of age-specific models of prevention"), the extension to other types of prevention is straightforward and fits into the scope of models presented in the present paper. Such an investigation is currently the main focus of concern.

Many issues concerning population dynamics include heterogeneity, either spatial or temporal. The age distribution of a population can change sometimes very quickly, and can be completely different in different regions. Hence, it is necessary that a model for adapting strategies to the different cases should incorporate the age distribution of population groups. The age-specific models presented in the present paper are a first step towards including more heterogeneity in drug epidemic models. Further extensions would be to use finer group classifications or to introduce the duration of use as a third parameter, because some aspects of drug initiation depend on how long an individual has been in his or her current group. For example, the probability of an individual moving from light to heavy use may

be related to the duration of his or her drug consumption. Theoretically, such an extension can be developed from the models discussed in the present paper, but, in the current state of science, the analysis of those models encounters difficulties on two sides: the availability of data for the large number of parameters and the numerical limitation due to computational power.

In summary, the age-specific models described in the present paper allow more detailed insight into the dynamics of drug epidemics. By formulating them as optimal control models, they can be used to increase the effectiveness of prevention programmes through improved targeting.

References

1. J. D. Murray, *Mathematical Biology* (Berlin, Springer-Verlag, 1993).
2. J. Kaplan, *The Hardest Drug: Heroin and Public Policy* (Chicago, University of Chicago Press, 1986).
3. M. A .R. Kleiman, "Enforcement swamping: a positive feedback mechanism in rates of illicit activity", *Mathematical and Computer Modeling,* vol. 17, 1993, pp. 67-95.
4. J. G. Bachman and others, "Explaining the recent decline in marijuana use: differentiating the effects of perceived risks, disapproval and general lifestyle factors", *Journal of Health and Social Behavior,* vol. 29, 1988, pp. 92-112.
5. S. N. Busenberg, M. Iannelli and H. R. Thieme, "Global behavior of an age-structured epidemic model", *SIAM Journal on Mathematical Analysis,* vol. 22, No. 4 (1991), pp. 1065-1080.
6. J. Müller, "Optimal vaccination patterns in age-structured populations", *SIAM Journal on Applied Mathematics,* vol. 59, No. 1 (1998), pp. 222-241.
7. S. S. Everingham and C. P. Rydell, *Modeling the Demand for Cocaine* (Santa Monica, California, RAND Corporation, 1994).
8. J. P. Caulkins and others, *The Cost-Effectiveness of School-Based Drug Prevention Programs* (Santa Monica, California, RAND Corporation, 1999).
9. G. W. Swan, *Applications of Optimal Control Theory in Biomedicine* (New York, Marcel Dekker, 1984).
10. B. L. Keyfitz and N. Keyfitz, "The McKendrick partial differential equation and its uses in epidemiology and population study", *Mathematical Computer Modelling,* vol. 26, No. 6 (1997), pp. 1-9.
11. D. F. Musto, *The American Disease* (New Haven, Connecticut, Yale University Press, 1999).
12. D. Behrens and others, "A dynamic model of drug initiation: implications for treatment and drug control", *Mathematical Biosciences,* vol. 159, 1999, pp. 1-20.
13. C. Almeder, G. Feichtinger and G. Tragler, "An age-specific light and heavy user model for illicit drug consumption", draft working paper (Vienna, Vienna University of Technology, 2000).
14. C. P. Rydell and S. S. Everingham, *Controlling Cocaine: Supply vs. Demand Control Programs* (Santa Monica, California, RAND Corporation, 1994).

15. B. D. Crane, A. R. Rivolo and G. C. Comfort, *An Empirical Examination of Counterdrug Interdiction Program Effectiveness* (Alexandria, Virginia, Institute for Defense Analysis, 1997).

16. D. Behrens and others, "Optimal control of drug epidemics: prevent and treat—but not at the same time?", *Management Science,* vol. 46, No. 3 (2000), pp. 333-347.

17. G. Tragler, J. P. Caulkins and G. Feichtinger, "Optimal dynamic allocation of treatment and enforcement in illicit drug control", research report 212 *Operations Research,* forthcoming.

18. A. Baveja and others, "When haste makes sense: cracking down on street markets for illicit drugs", *Socio-Economic Planning Sciences,* vol. 31, No. 4 (1997), pp. 293-306.

19. A. V. Naik and others, "Scheduling crackdowns on illicit drug markets", *European Journal of Operational Research,* vol. 88, 1996, pp. 231-250.

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

The epidemiology of drug use at the macro level: indicators, models and policy-making*

L. WIESSING
R. HARTNOLL
European Monitoring Centre for Drugs and Drug Addiction, Lisbon, Portugal

C. ROSSI
Department of Mathematics, University of Rome "Tor Vergata", Rome, Italy

ABSTRACT

Indicators and models may help in studying and understanding the epidemiology of drug use at the international or macro level. Indicators of drug use are usually based on available routine data and give indirect information on prevalence and trends. They can be used to cover large geographical areas with a limited budget; however, they are usually limited in quality and scope. Modelling is based on mathematical theory and can be used to integrate data from different indicators and other sources. It can be used to estimate prevalence and incidence or to enhance the understanding of drug processes by simulating experiments that are difficult or impossible to perform in real life. A sufficient understanding of basic assumptions and limitations is crucial to the interpretation of modelling results. Depending on the availability and use of data, a whole continuum of modelling may exist, ranging from empirical analyses with a high applicability to real life, to theoretical hypothesis-generating exercises based on many assumptions. Compared with other areas, the epidemiology of drug use has so far made little use of modelling, partly because reliable data remain scarce; nevertheless, work on improving indicators and other data is in progress. In the field of drugs, some issues that continue to hamper the application of evidence-based policy-making are lack of a common understanding of priorities for study and intervention, lack of knowledge of the spread and progression of drug problems, and legal and moral discourses. Indicators and models, which are important, complementary tools for epidemiology at the macro level, may help to clarify such issues. Although the quality of data and thus the inferences to be drawn from them are often more limited than they are in in-depth empirical studies, macro-level assessments are necessary to guide national and international policy decisions.

The present article benefited from the European Network to Develop Policy-Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions, funded by the European Union (DG Research, Targeted Socio-Economic Research project ERB 4141 PL 980030). An earlier version of the article has appeared in European Monitoring Centre for Drugs and Drug Addiction scientific monograph number 6: *Modelling Drug Use: Methods to Quantify and Understand Hidden Processes* (Lisbon, European Monitoring Centre for Drugs and Drug Addiction, 2001). The authors are grateful for the helpful comments of Alred Uhl.

Introduction

One problem of drug epidemiology at the national and international, or macro, level is the sheer scale of activities. While data on drug use within a city or small region can be collected directly, usually no funding is available for studies covering large areas or a number of countries. Setting up such a study is not straightforward, because of differences between countries and regions and the many stakeholders involved. Data available at the macro level are often aggregated rather than describing individual cases; they are based on routine sources, with their limitations of analysis and interpretation. The question that arises is whether valid inferences can still be made at the macro level (and, if so, how). The present article aims to introduce briefly indicators and modelling, two different but complementary tools for use in policy-making at the macro level, and to point to some still unresolved problems in the field of drugs that interfere with such evidence-based approaches.

Indicators

Epidemiologists describe the spread of disease in a population in order to provide evidence for public health interventions and policies. Basic epidemiological measures are prevalence, referring to all existing cases at a certain moment in time, and incidence, referring to all newly occurring cases in a certain time period. In order to understand causal factors that lead to disease, a risk factor analysis is usually performed by comparing individual cases with non-cases. In drug epidemiology, it is usually difficult to determine prevalence and incidence owing to the hidden nature of drug use. Normal sources of data such as medical services may be very incomplete, while standard tools such as household surveys may give biased response owing to social stigma and the low rate of social integration of heavy drug users. Added to those problems is the difficulty of collecting data of comparable quality from a large number of sources and countries when working at the macro level. Additional methods are necessary to estimate prevalence, incidence and trends in drug use over time.

In domains such as public health and economy, the concept of indicators has been developed for data collection on a large scale and for phenomena that are difficult to measure. Rather than aiming at exact prevalence and incidence figures, indicators are based on available routine data on disease, risk factors or consequences and may provide indirect information on the exact magnitude of and trends in the disease. Public health indicators rely on data sources such as mortality registers, population censuses, routine health-service records, epidemiological surveillance data, sample surveys or disease registers. The idea is to select variables from such sources that fulfil certain quality requirements (validity, reliability), in order to measure health and changes in health [1].

Work on indicators in the field of drug use began in London in the early 1970s. It spread across European cities via the Pompidou Group of the Council of Europe, continued to be developed in countries across Europe by the European

Monitoring Centre for Drugs and Drug Addiction (EMCDDA) and is currently being implemented at the global level by the United Nations International Drug Control Programme [2-4]. Available data on drug use are limited and the choice of indicators has relied on a pragmatic approach. Currently, five so-called key indicators are being implemented by EMCDDA in the European Union: *(a)* general population and school surveys; *(b)* estimates of "problem" drug use; *(c)* data from drug treatment services; *(d)* drug-related deaths and mortality; and *(e)* drug-related infections (such as the human immunodeficiency virus (HIV) and hepatitis B and C) [4]. Standards have been developed and current work is focused on collecting data and overcoming practical problems in data quality and comparability. Other indicators are still at an early stage of development, such as those related to social problems and crime, hospital and emergency room data and data from popular meeting places of young people, including discos and nightclubs.

Interpreting data based on drug indicators is not always straightforward. It is sometimes assumed that the combined information of a set of indirect indicators reflects trends in the prevalence of problematic drug use; however, indicator data may show fluctuations unrelated to prevalence due to, for example, changes in heroin supply or decreases in risk behaviour. Indicators such as drug-treatment data, drug-related mortality and drug-related infectious diseases are, therefore, also important for their own sake and for evaluating interventions directed at each of the specific sub-populations at risk, in addition to being important for following trends in problematic drug use.

Indicator data are often of limited quality owing to their wide-ranging coverage and routine nature; thus it is dangerous to rely solely on indicators. Most routine data are not suitable for providing rapid insight into changing trends in problematic drug use. For instance, the average latency time of 5-8 years between first heroin use and first demand for treatment implies that treatment indicators are not suitable for monitoring drug use in the non-treated population [5, 6]. To interpret trends at the national level, such as an increase in drug-related deaths, and in order for appropriate measures to be taken, indicators should be complemented and validated by smaller in-depth studies. Regularly repeated, small-scale studies are also important in giving insights into possible or changing causes, risk factors or confounding factors of the observed trends, as additional background information can usually be gathered from such studies. Routine indicators and one-off or repeated studies together may give a more complete picture of a phenomenon such as problematic drug use. Indicators provide lower-quality data but at low cost and with wide geographical coverage, while local studies provide high-quality data from a more limited and often high-risk area.

Dynamic models

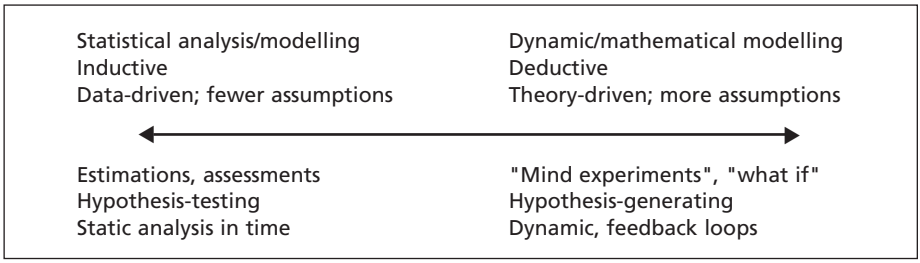
Dynamic modelling originated from biology and infectious disease epidemiology [7]. Many recent advances in research were made in the field of infectious diseases, especially on the acquired immunodeficiency syndrome (AIDS) [8]. Over the last 15 years, developments in modelling have closely paralleled the increasing

knowledge in etiology and transmission of AIDS and other sexually transmitted diseases. During that period, modelling studies moved from straightforward extrapolations of available AIDS data to highly complex transmission models with large sets of parameters. Modelling and empirical research on AIDS have been mutually informative about important concepts. While characteristics of patients pointed to the existence of specific groups with increased risk, such as homosexual men and injecting drug users, dynamic modelling clarified the potential role of such "core groups" (with high-risk behaviour or assortative "like-with-like" mixing patterns and consequent high levels of infection) for the overall transmission dynamics within a susceptible population [9]. While the natural history of HIV infection was becoming clearer from cohort studies, modelling offered a means of clarification and simplification (for example, by determining a number of discrete disease stages), thus providing a basis for policy decisions (for example, on treatment needs per disease stage) [10]. The effect of potentially important biological parameters that could not easily be studied empirically were often postulated from dynamic modelling work. For example, a simulation model could show that the high level of infectiousness during the first weeks of HIV infection could in some cases determine the occurrence of most new infections in the population, raising some questions about the effectiveness of HIV screening as a prevention tool under all circumstances [11, 12].

For drug epidemiology at the macro level, dynamic modelling can be a valuable complementary tool for following trends in indicators and other direct data analysis. In the more usual inductive or empirical method of data collection and interpretation, new insights follow from observation. Conversely, dynamic modelling may often be nearer a deductive approach, where new insights follow from theory. The use of the outcomes in real-life situations is sometimes limited and it can be difficult to validate the model assumptions with the data. However, the importance of most types of dynamic modelling lies in the generation of theory and hypotheses that may provide a framework for further research and policy decisions.

It is not easy to define dynamic or mathematical models. In order to understand how they are used, it may be helpful to compare them with the more familiar statistical models, used for empirical data analysis, although even there the difference is not sharp (see figure I). Most models consist of a set of mathematical equations (often in the form of a computer program) that describe a process (such as the spread of drug use) in a simplified manner and that can be used to understand the behaviour of that process. A whole continuum of models exists, between the extremes of pure empirical data analysis with statistical models and using mathematical or economic theory in dynamic models. Statistical modelling techniques, such as different types of regression analysis, including factor, cluster and path analysis, often rely much more on data than, for example, many compartmental or system dynamics models, which are more theoretical. Such dynamic models may rely more heavily on sometimes unproven assumptions and may often be based on "thin" data (data of which the quality is not certain), which makes their validity more difficult to determine.

Figure I. Some differences between statistical and dynamic modelling



Note: This scheme does not fit all models. There are mathematical models that are data-driven and statistical models that are based on very little data. In addition, some statistical analysis (such as time series analysis and survival analysis) is not static in time and deterministic mathematical models can be used for estimations. Many models, however, appear to fit this scheme.

Both extremes of the modelling continuum are useful in macro-level epidemiology, but in different ways. Rigorous analysis of statistical data, such as regression analysis based on individual data records, is usually not possible at the macro level, but statistical models can often suggest useful answers when there are gaps in data of a certain type. That can happen, for example, by imputing (that is, interpolating) prevalence data from other moments in time or locations where more data are available, or by using other data sources and if possible adjusting for potential biases, making it possible to generate a result where originally no data were available (although again special care must be taken in interpretation). There is a clear need for such modelling in the field of drugs at the macro level, where reliable data often do not exist.

At the other end of the continuum, dynamic or mathematical models can assist in understanding a phenomenon even if such models are almost entirely based on assumptions, with little or no data input. If the model is thought to describe a process in real life sufficiently well, the behaviour of such a process can be studied under different circumstances by varying parameter values and observing the variation in outcomes. Such a semi-experimental situation is sometimes called "what if" modelling, that is, the model explores what could happen under certain circumstances, if it can be assumed that the process under study has been captured by the model. Modelling, therefore, can provide a tool for simulating experiments that are not possible to perform in real life for practical or ethical reasons.

To interpret the results of modelling, it is important to understand the differences between the different types of model. Results can be seen either as scientific "facts" (statistical modelling), as theoretical hypotheses or even as something in-between, depending on the certainty of parameters and the validity of the basic model chosen. The limitations and assumptions of the model should always be stated clearly and presented together with the results. That is often difficult, even for specialists, and non-specialists may have no way of distinguishing valid results with important direct implications from academically interesting theory with little direct practical meaning.

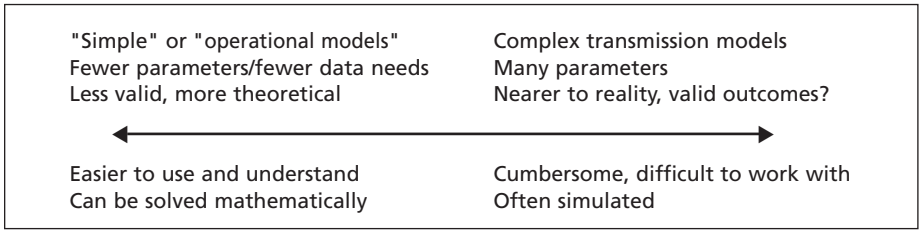
To avoid such problems, it is important for modellers and other scientists to form multidisciplinary teams, although that is possible only if each team member has some understanding of the approaches used in the other disciplines and is prepared to acquire at least some knowledge of other fields. If such collaboration exists, the modeller may generate meaning from scarce data and the topic expert, such as the epidemiologist, can help to interpret the validity of the results against his or her knowledge of real-life situations and data quality issues.

If one concentrates only on dynamic or mathematical modelling (but most of the following also holds for statistical analysis/modelling), again different approaches are possible (see figure II). A relatively simple model that reflects the main elements and their relationships is often the starting point. A common criticism of such a model is that it fails to take into account a potentially important variable, which may subsequently be incorporated into the model, making the model more complex. The process may repeat itself until it results in highly complex models. Although the more complex models may resemble reality more closely than the simple ones do, it is often difficult to keep track of what exactly is happening. In practice, the complex models may not lead to better or more valid results. As was the case with regard to AIDS research, the process of increasing model complexity often directly reflects the state of existing knowledge in the field. However, while some modellers attempt to incorporate into their models as many parameters as possible in an attempt to increase the validity of their results, others prefer to work with simple models that can easily be understood. Simple models that succeed in giving new insights are often more influential than complex ones [13].

Sensitivity analysis is an important element, in particular in the case of more complex models. In sensitivity analysis, the relative influence of the different parameters on the modelling results is assessed. Different, often extreme, values are taken for those parameters. It may be found that the results are not sensitive to some of the parameters but very sensitive to others, making it necessary for those parameters to be especially reliable. In addition, the results of a complex modelling study can be analysed by applying statistical sampling techniques to the modelling results, as shown in a study of the spread of HIV in New York [14]

Another important question during the interpretation of modelling results is their intended use. Depending on the application of modelling results, one should be more or less conservative. A set of outcomes can be valid for publication in a reputable mathematical journal, where the emphasis is on methodological and

Figure II. Two basic approaches to dynamic modelling



theoretical advances, while the same outcomes may not be considered sufficiently rigorous to be published in a reputable applied journal (for example, one on public health), that focuses on the external validity of results and their practical consequences. Nevertheless, even for applied use, the question is sometimes whether it is better to have less valid information than no information at all, which may lead to results being accepted for policy decisions that are less valid than would be desirable from a strictly scientific point of view.

An example of the latter situation in drug epidemiology is prevalence estimation. Although sophisticated statistical models exist for prevalence estimation of problematic drug use at the local level, such as multi-sample capture-recapture models with loglinear regression parameter estimation [15], results may often not be clear owing to problems in data quality: they often have wide confidence intervals, which may still substantially underestimate uncertainty. At the macro level, such methods can often not easily be used because of geographical heterogeneity and the limited availability of national data. Simple multipliers may be used instead, while in some cases the uncertainty about data quality may make calculating confidence intervals meaningless [16]. At best, these methods give a qualitative indication of a range for plausibility of prevalence [17]. Nevertheless, it is important to have prevalence estimates at the local and national levels that form a basis for a large variety of policy choices. In general, low-quality estimates are accepted as the best approximations currently available. Other examples of accepting uncertain modelling results can be found in forecasts of future developments and in scenario analyses, which by definition are unreliable and may eventually prove to be wrong, but are of such importance to policy makers that they are widely used.

Policy needs and problems using indicators and models

Epidemiology is an applied or problem-driven science in which emphasis is placed on the relevance of new knowledge to policy. It may therefore be useful to discuss briefly the possible needs of policy makers. This is especially important at the macro level, where the relevance of policy for individual drug users may be much less obvious than, for example, at the level of treatment or outreach services.

The needs of a policy maker may seem clear in general terms: it is necessary to describe and understand a problem and follow trends, to design appropriate interventions and to evaluate the results of interventions. In the case of drug use and its consequences, both indicators and models may contribute to a description and understanding of the problem, while indicators are specifically suited to following trends. Models may help in designing and choosing interventions, for example, by assessing the cost-effectiveness of alternative options, although the basic knowledge about what works (that is, the efficacy of different options) needs to be assessed in carefully designed empirical studies.

The reality, however, is less harmoniously arrived at than the above would suggest. At a global level, there is little consensus among scientists, service providers

and policy makers about the most effective and relevant interventions. Opinions range from a "war on drugs" at one extreme to complete legalization at the other. The question that arises is what makes the field of drugs seem so much more complex than it appears to be in the first instance. In part, the answer may be of a purely scientific and methodological nature; however, it may also lie in part in the interplay between science and policy. Furthermore, some of the more methodological problems may be a consequence of working at a high level of aggregation.

Scientific or methodological problems, especially at the macro level, may include the lack of a clear case definition of what constitutes a drug user, the lack of knowledge of basic mechanisms for the spread of drug use and the lack of knowledge regarding progression into drug problems. These three areas are discussed below.

Firstly, using indicators at the macro level, it is often not possible to use a clear case definition of addiction or drug dependence, such as the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* [18], or to use standard disease codes, such as the codes of the International Classification of Diseases, because of the quality and limitations of routine data. Indicators such as survey data may give data of high quality on lighter forms of drug use, but are expensive and usually not as reliable for heavier patterns of drug use, which cause most drug problems. Modelling techniques for prevalence estimation of problem drug use, such as capture-recapture, often rely on non-specialized data sources that give little detailed information on drug use (police data, hospital data, etc.) and make clear-cut case definitions inappropriate. In such cases, a pragmatic solution may be to define problem drug users as drug users who are in contact or in need of contact with health or social services.* In most countries in Europe, such a definition is usually limited to frequent heroin or amphetamine users; however, it is difficult to obtain more detail on the prevalence and patterns of problematic drug use, such as breakdowns by type of drug, from those techniques and it is important that they are complemented by local studies, based on, for example, out-of-treatment recruitment of drug users.

Secondly, more knowledge is required about the basic mechanisms for the spread of drug use and problem drug use or addiction. There is little consensus among scientists on which risk factors are the most influential, from the wide range of social, psychosocial and biological factors identified in different studies [19]. Until that situation changes, effective interventions cannot be expected. Such a lack of common understanding about a basic mechanism for the epidemiology of drug use also makes it difficult for consensus to be reached on results from modelling studies, which is not the case in more established fields such as infectious disease epidemiology or economy. Although models of spread are often partly based on assumptions, they can help to clarify the underlying processes. For example, there is evidence that drug use spreads like an infectious disease, that is, the rate of new cases depends on the number of existing cases and the number of

*Although EMCDDA defines problem drug use as "injecting drug use (IDU) or long-duration/regular use of opiates, cocaine and/or amphetamines", the estimates are in practice calculated from data on users contacting services.

those that are susceptible [20, 21]. Using the analogy with infectious diseases enables a large body of existing research to be used for the study of the spread of drug use. However, unlike the case of infectious diseases, the role of supply factors (including price of drugs and availability) is not well understood nor are the intentional marketing activities of drug dealers. In addition, "infectiousness" (that is, the probability of "infecting" another person) may not be as constant as in a biological context and may depend on a range of unknown social factors. For example, the spread of drug use may be influenced by the mass media, through developments in the music and fashion industries. Such factors may have contributed to the recent spread of Ecstasy (methylenedioxymethamphetamine (MDMA)) and other new synthetic drugs.

Thirdly, more work is needed to clarify the "natural history" or typical course of progression to problematic drug use. It may be less important to study what leads young people to experiment with illegal drugs, as that mechanism does not seem to differ from what leads people to experiment with alcohol or tobacco at earlier ages [22, 23]. It is, however, crucial to understand why some young people progress into using heroin dependence, while by far most never go beyond experimenting with cannabis. This may be due simply to social factors such as "meeting the wrong friends". It may also be that some people are genetically prone to addiction or have other predisposing factors and that those who continue using drugs need to do so to ease negative moods or to "self-medicate" mental problems [24]. In the first case, again, a type of modelling similar to that for infectious diseases could be useful in, for example, predicting progression to heroin dependence and estimating cost-effectiveness of interventions. Concepts from infectious disease epidemiology might be applied, such as "infecteds", "susceptibles", "basic reproductive rate"* and "herd immunity"**. Key issues would be to distinguish "susceptibles" from "immunes" or to find a type of "vaccination" that could turn "susceptibles" into "immunes". Prevention measures to be considered might then resemble those used in the infectious diseases field. In the second case, the epidemiology of illegal drug use might resemble more closely that of chronic diseases. The models would focus on demographic or social developments that would lead to higher or lower prevalence and incidence of the predisposing genetic or mental-health factors in the general population. It would then be important to identify those with a genetic or mental risk profile at an early stage and to find ways to protect such individuals from addiction to a substance or to a behaviour (such as gambling) later in life. In fact it is probable that a mix of both types of approaches should be used, that is, experimentation with drugs might follow infectious diseases patterns, while continued heavier use (dependence) may be more related to chronic diseases mechanisms.

In addition to the examples of scientific and methodological problems described above, other problems in the study of drug use exist, illustrating the intersection of science and policy. Possibly the clearest example is the illegal

*"Basic reproductive rate": average number of new cases "infected" by each existing case.

**"Herd immunity": protective effect of a high density of immune persons on the non-immunes in the population.

status of drug use, which may affect several problems in the field, both for scientists and policy makers. There are few epidemiological areas in which the disease under study is viewed as an illegal activity. The legal and moral aspects may complicate public health work on drug problems in the ways described below.

Firstly, methodologically sound studies on interventions are difficult to set up, because of the illegal status of patients and because some of the potential interventions are illegal. Those factors complicate the collection of hard evidence on policy options; thus, the debate continues in a vicious circle. An example of a controversial intervention is controlled heroin provision [25, 26]. Even substitution treatment, an alternative intervention that is well accepted in the scientific field [27], cannot be provided everywhere. Patients treated in law enforcement settings such as prisons do not always have the possibility of a confidential doctor-patient relationship. In addition, patients may be moving in and out of treatment and prisons, thus continuously interrupting therapy, which makes positive effects difficult to sustain. Such factors also greatly increase the risk of adverse consequences, including HIV infection or overdose [28-31]. Moreover, the legal restrictions and stigma concerning drug use in most countries may lead to difficulties in obtaining reliable routine data or indicators on the problem [32]. Drug users tend to avoid registries as much as possible, and some doctors have been known to give a drug-related death a less stigmatized code on the death certificate in order to protect the relatives of the deceased. Legal restrictions may affect scientific work in a more general manner, through their influence on funding priorities or pressure on scientists to select results that are politically correct.

Secondly, legal and moral perspectives may lead to a lack of clarity about the case definition in policy debates. Illegal drug use is often discussed only in general and vague terms; however, it is clearly important to distinguish problem drug use from non-problem drug use or cannabis use from heroin. In many countries cannabis use (which is usually non-problematic) does not differ in legal terms from heroin use (which is often problematic), yet the consequences in terms of public health are vastly different. Many forms of drug use, even if illegal, do not lead to more problems for the individual or society than does moderate smoking or alcohol use [33, 34]. Problematic drug use, such as injecting or frequent use of heroin, crack and other hard drugs, incurs most costs to society in the form of infections, deaths, addiction and crime [4, 35]. The aim of a policy maker is to try to minimize the social costs of drug use by formulating and implementing appropriate policy adapted to such an aim. It is thus those forms of drug use which incur the largest social costs which should be targeted first by interventions. Drug policies that do not distinguish between different modes of drug use may not be efficient and may even be counter-productive. Differentiating between light and heavy forms of drug use in policy objectives, and using case definitions based on clear diagnostic rather than legal criteria, could contribute substantially to progress in the field of drugs.

Thirdly, the lack of clarity in the definition of the problem and the lack of possibilities to gather scientific evidence have resulted in a lack of common objectives to reduce drug problems. Whereas there is no question about whether HIV

infection or cancer should be prevented or not, such a consensus does not exist with regard to drug use. It is important to reach evidence-based agreement on general objectives and on an optimal balance of intervention priorities to reach those objectives [36]. Some decision makers believe that reducing all drug use (if it is possible to do so by law enforcement) will in turn lead to less problematic use, while others think that it is not possible and that law enforcement actually increases problematic use, having little or no effect on the prevalence of non-problematic drug use [37]. Both approaches, however, imply completely different and often conflicting interventions: legal and preventive in one and oriented towards public health in the other. It has been found that changes in law enforcement usually have little effect on levels of problematic drug use [38, 39] and that countries with totally different policies may show similar prevalence [4]. For those reasons, it may be more useful to aim at preventing problematic drug use and its secondary consequences, such as overdose and HIV transmission [40, 41], than to attempt to eliminate or prevent all forms of drug use. Most policies regarding alcohol abuse also focus on preventing problematic alcohol use and its consequences than on attempting to eliminate alcohol consumption altogether.

Conclusion

Five key indicators are being implemented in the member States of the European Union while others are still being developed. Data collection has started only recently. The quality of data and comparability issues have not yet been resolved. Many models of drug use exist but consensus on the basic mechanisms for the spread and progression of problematic drug use is still lacking, while legal issues surrounding drug use significantly complicate data collection and monitoring of problematic drug use, as well as the development and evaluation of effective interventions.

It is essential to improve the quality of the monitoring of problematic drug use. While non-problematic drug use can be followed through general population and school surveys, prevalence estimation of problematic drug use is still in its infancy. Prevalence and incidence are fundamental measures aimed at understanding the spread of problematic drug use and evaluating the effects of policy interventions. While prevalence by subgroups can provide information on the exit rates from the population (for example, through treatment), detailed information on incidence is important for assessing entry rates and evaluating preventive measures.

At present, the quality of the data does not permit the precise measurement of prevalence and incidence in most countries, even though important methodological advances are being made. It is necessary to obtain a greater commitment to quality data collection for both statistical estimation modelling and developing dynamic models in order to investigate policy choices. Legal barriers (including data protection laws) have to be removed to facilitate scientific work on all potentially useful interventions and to improve data availability.

More multidisciplinary work is needed to clarify and describe the spread of problematic drug use in space and time. Economic models need to be developed to enable better estimates to be made of the costs to society of the various consequences of drug use and of the cost-effectiveness of policy options. There is a need to clarify the role of supply factors, as well as the ways in which those factors interact with demand research into drug markets and, in particular, with progression to problematic drug use. There is a more advanced understanding of the spread of drug-related infectious diseases; however, practical work is needed on HIV prevention (such as increasing the coverage of interventions), and the prevention of hepatitis C infection still needs to be enhanced by improving injecting drug users' understanding of the routes of transmission of those viruses. As effectiveness of interventions cannot be assessed easily at the macro level, monitoring of the disease should at least include monitoring of the coverage of the target population [42].

In conclusion, indicators and models constitute important and complementary tools for the epidemiology of drug use at the macro level. Although the quality of the data and the inferences drawn from them are often weaker than those in local or in-depth studies, macro-level assessments are indispensable for guiding policy decisions. The interpretation of results should be carefully undertaken at the macro level, despite the urgent need for policy decisions to be made. There are still many problems specific to the field of drugs that need to be solved before effective measures can be widely adopted.

Annex

European Network to Develop Policy Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions

A European network was established in 1999 with the aim of stimulating the use of modelling in the field of drug use. It is funded by the European Commission (DG Research, Targeted Socio-Economic Research) and coordinated by the European Monitoring Centre for Drugs and Drug Addiction. Six working groups cover three broad areas: two working groups are active in the field of prevalence estimation (at the local level and the national level), two are investigating the dynamics of drug use (geographical spread, time trends and incidence) and two are studying the economic aspects of drug use (costs, cost-effectiveness and drug markets). About 40 European modellers and other experts are participating in these working groups. Several advances have been made in the two years that regular meetings have been held. In the area of prevalence estimation, new methods (such as the truncated Poisson method) have been studied that may provide prevalence estimates from fewer data sources than are necessary in classical capture-recapture and methods that may account for non-closed populations. At the national level, a multivariate method has been applied that may be more robust than using simple multipliers (the multivariate indicator method). A model of geographical spread has been developed and linked with a geographic information system application that can be used to visualize spread between cities and

smaller towns, based on various parameters, including geographical data on the year of peak prevalence and population size. A method has been developed to estimate incidence (initiation rates) of problematic drug use from treatment data using a back-calculation model analogous to the one used for AIDS. That work has also resulted in estimates of the latency time between first heroin use and first treatment for opiate addiction, in different cities in Europe, which is important for understanding the natural history of progression to heroin use. Different methods have also been developed to estimate the social costs of problematic drug use, placing emphasis on estimating costs of drug-related infections (notably HIV and hepatitis B and C); drug markets and supply-side factors have also been studied. The collaboration of many experts during the project has led to the publishing of a substantial amount of scientific work on the epidemiology of drug use in Europe. (For further details and a list of publications, see the information on the Internet at http://www.emcdda.org/situation/methods_tools/modelling_ network.shtml)

References

1. World Health Organization, *Teaching Health Statistics* (Geneva, 1999).
2. R. Hartnoll, "Epidemiological approaches to drug misuse in Britain", *Journal on Addictive Diseases,* vol. 11, No. 1 (1991), pp. 33-45.
3. R. Hartnoll and others, "A multi-city study of drug misuse in Europe", *Bulletin on Narcotics* (United Nations publication), vol. 41, Nos. 1-2 (1989), pp. 3-27.
4. European Monitoring Centre for Drugs and Drug Addiction, *2000 Annual Report on the State of the Drugs Problem in the European Union* (Lisbon, European Monitoring Centre for Drugs and Drug Addiction, 2000).
5. C. Rossi, "Monitoring drug control strategies: hidden phenomena, observable events, observable times", *International Journal of Drug Policy,* vol. 10, 1999, pp. 131-144.
6. L. G. Wiessing and others, "Analysis of the latency period from onset of heroin use to first treatment; comparisons among four European sites, possible implications for treatment services and for understanding the natural history of problematic heroin use", unpublished.
7. R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford, Oxford University Press, 1992).
8. C. Pasqualucci, L. Ravà, C. Rossi and G. Schinaia, "Estimating the size of the HIV/AIDS epidemic: complementary use of the Empirical Bayesian Back-Calculation and Mover-Stayer model for gathering the largest amount of information", *Simulation* (San Diego, California), vol. 71, 1998, pp. 213-227.
9. M. Kretzschmar, Y. T. van Duynhoven and A. J. Severijnen, "Modeling prevention strategies for gonorrhea and chlamydia using stochaastic network simulations", *American Journal of Epidemiology,* vol. 144, No. 3 (August 1996), pp. 306-317.
10. I. M. Longini and others, "Statistical analysis of the stages of HIV infection using a Markov model", *Statistics in Medicine* (Chichester, United Kingdom), vol. 8, No. 7 (July 1989), pp. 831-843.

11. E. S. Daar and others, "Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection", *New England Journal of Medicine* (Boston), vol. 324, No. 14 (4 April 1991), pp. 961-964.

12. M. Kretzschmar and L. G. Wiessing, "Modelling the spread of HIV in social networks of injecting drug users", *AIDS* (London), vol. 12, No. 7 (7 May 1998), pp. 801-811.

13. E. H. Kaplan and E. O'Keefe, "Needles that kill: modeling HIV transmission via shared drug injection equipment in shooting galleries", *Review of Infectious Diseases* (Chicago), vol. 11, 1989, pp. 289-298.

14. S. M. Blower and others, "Drugs, sex and HIV: a mathematical model for New York City", *Philosophical Transactions of the Royal Society of London, Series B Biological Sciences* (London), vol. 331, No. 1260 (28 February 1991), pp. 171-187.

15. M. Frischer and others, "Estimating the population prevalence of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland", *American Journal of Epidemiology,* vol. 138, No. 3 (1 August 1993), pp. 170-181.

16. European Monitoring Centre for Drugs and Drug Addiction, "Study to obtain comparable national estimates of problem drug use prevalence for all EU member states", 1999.

17. D. Seidler and A. Uhl, *Prevalence Estimate of Problematic Opiate Consumption in Austria* (Vienna, LBISucht, 2000).

18. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders,* 4th ed. (Washington, D.C., American Psychiatric Association, 1994).

19. T. Rhodes and others, "Risk factors associated with drug use: the importance of 'risk environment'", 1999.

20. D. R. Mackintosh and G. T. Stewart, "A mathematical model of a heroin epidemic: implications for control policies", *Journal of Epidemiology and Community Health* (London), vol. 33, 1979, pp. 299-304.

21. R. de Alarcón, "The spread of heroin abuse in a community", *Bulletin on Narcotics* (United Nations publication), vol. 21, 1969, pp. 17-22.

22. D. Kandel and R. Faust, "Sequence and stages in patterns of adolescent drug use", *Archives of General Psychiatry* (Chicago, vol. 32, 1975, pp. 923-932.

23. A. Bauman and P. Phongsavan, "Epidemiology of substance abuse in adolescence: prevalence, trends and policy implications", *Drug and Alcohol Dependence* (Coclare, Ireland), vol. 55, 1999, pp. 187-207.

24. A. I. Leshner, "Science based views of drug addiction and its treatment", *Journal of the American Medical Association,* vol. 282, 1999, pp. 1314-1316.

25. W. Van den Brink, V. M. Hendriks and J. M. van Ree, "Medical co-prescription of heroin to chronic, treatment-resistant methadone patients in the Netherlands", *Journal of Drug Issues* (Tallahassee, Florida), vol. 29, 1999, pp. 587-608.

26. A. Uchtenhagen and others, *Prescription of Narcotics for Heroin Addicts: Main Results of the Swiss National Cohort Study* (Basel, Karger, 1999).

27. D. C. Des Jarlais and R. Hubbard, "Treatment for drug dependence", *Proceedings of the Association of American Physicians* (Philadelphia), vol. 111, 1999, pp. 126-130.

28. D. S. Metzger, H. Navaline and G. E. Woody, "Drug abuse treatment as AIDS prevention", *Public Health Reports* (Washington, D.C.), vol. 113 (suppl. 2), 1998, pp. 97-106.

29. P. Turnbull, "Prisons: Vancouver Conference Review", *AIDS Care* (Abingdon, United Kingdom), vol. 9, 1997, pp. 92-95.

30. L. G. Wiessing and others, "Prevalence and risk factors for HIV infection among drug users in Rotterdam, the Netherlands", XIth International Conference on AIDS, Vancouver, Canada, 7-11 July 1996.

31. E. J. van Ameijden, M. W. Langendam and R. A. Coutinho, "Dose-effect relationship between overdose mortality and prescribed methadone dosage in low-threshold maintenance programs", *Addictive Behaviors* (Oxford), vol. 24, 1999, pp. 559-563.

32. P. H. Hughes and O. Rieche, "Heroin epidemics revisited", *Epidemiologic Reviews* (Baltimore, Maryland), vol. 17, 1995, pp. 66-73.

33. T. Nicholson, J. White and D. Duncan, "A survey of adult recreational drug use via the World Wide Web: the Drugnet study", *Journal of Psychoactive Drugs* (San Francisco), vol. 31, 1999, pp. 415-422.

34. C. Marwick, "Illicit drug users not idle: report says 70% go to work", J*ournal of the American Medical Association,* vol. 282, 1999, pp. 1320-1321.

35. H. Amaro, "An expensive policy: the impact of inadequate funding for substance abuse treatment", editorial, *American Journal of Public Health* (New York), vol. 89, 1999, pp. 657-659.

36. R. Hartnoll, "A framework for evaluating activities to reduce the demand for drugs in Europe", *GACETA Sanitaria* (Barcelona), vol. 4, No. 21 (November 1990), pp. 239-251.

37. A. Uhl, "Non-problematic use—problematic use: a paradox", in *Illicit Drugs: Patterns of Use—Patterns of Response, Proceedings of the 10th Annual ESSD Conference on Drug Use and Drug Policy in Europe,* A. Springer and A. Uhl, eds. (Innsbruck, Studienverlag, 2000).

38. R. MacCoun and P. Reuter, "Interpreting Dutch cannabis policy: reasoning by analogy in the legalization debate", *Science* (Cambridge, Massachusetts), vol. 278, 1997, p. 47-52.

39. L. Lenke and O. Olsson, "Sweden: zero tolerance wins the argument?" *European Drug Policies and Enforcement,* N. Dorn, J. Jepsen and E. Savona, eds. (London, Macmillan Press, 1996).

40. G. V. Stimson, "Has the United Kingdom averted an epidemic of HIV-1 infection among drug injectors?" *Addiction* (Abingdon, United Kingdom), vol. 91, No. 8 (August 1999), pp. 1085-1088.

41. A. Wodak, "Harm reduction: Australia as a case study", *Bulletin of the New York Academy of Medicine,* vol. 72, 1995, pp. 339-347.

42. J. P. Habicht, C. G. Victora and J. P. Vaughan, "Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact", *International Journal of Epidemiology* (Oxford, United Kingdom), vol. 28, 1999, pp. 10-18.

Offprint from Bulletin on Narcotics, vol. LIII, Nos 1 and 2, 2001

Incidence indicators for policy-making: models, estimation and implications*

L. RAVÀ
Department of Mathematics, University of Rome "Tor Vergata", Italy

M. G. CALVANI
Department of Mathematics, University of Rome "Tor Vergata", Italy

S. HEISTERKAMP
Department of Clinical Epidemiology and Biostatistics, University of Amsterdam, the Netherlands

L. WIESSING
European Monitoring Centre for Drugs and Drug Addiction, Lisbon, Portugal

C. ROSSI
Department of Mathematics, University of Rome "Tor Vergata", Italy

ABSTRACT

Modelling the incidence of first opiate use can be a useful tool for understanding the process of diffusion of drug use in space and time. Such information is important in evaluating current and future needs for and effects of services and interventions.

Back calculation can be used in the analysis of spatial and time trends of incidence of drug users who will eventually seek treatment. This is done using incidence data of drug users entering first treatment and an external estimate of the latency period distribution, that is, the time lapse between first use and first treatment. The authors of the present article studied the heroin epidemic in Italy using Empirical Bayesian Back Calculation, at the national level and for eight regions of the country, using data from public treatment services for the period 1986-1998, which are routinely collected by the Ministry of Health. After a period of rapid increase during the 1980s, the nationwide aggregated incidence curve seemed to stabilize at the beginning of the 1990s. While that trend might possibly indicate some success in prevention efforts, it could well be due to the fact that the national epidemic is the combination of several local epidemics that emerge, saturate and move geographically. That hypothesis is confirmed by the regional analysis, which provides a picture of several sub-epidemics showing peaks of different size, location and time. In particular, Liguria, Piedmont and Emilia Romagna peaked more or less simultaneously before the other regions, followed by Tuscany, Latium and Sardinia (which show the highest incidence per million inhabitants), while Trentino and Friuli never peaked.

*The work described in the present article benefited from the European Union-funded European Network to Develop Policy Relevant Models and Socio-Economic Analyses of Drug Use, Consequences and Interventions. The authors thank Dr. Giovanni Nicoletti of the Ministry of Health for providing national drug treatment data for Italy.

The results of the Empirical Bayesian Back Calculation form an important evidence base for planning, monitoring and evaluating preventive interventions and for forecasting treatment needs in the near future, at both the regional and the national level.

Introduction

The implementation of drug policy needs to be evaluated in order to set up more efficient strategies to control drug-related phenomena and to forecast future service needs.

To evaluate interventions aimed at prevention and care for drug users, suitable indicators are necessary. In particular, epidemiological indicators may provide some evidence of the effectiveness of interventions directed towards drug users (control intervention) or the population at risk (prevention intervention). The possible prevention and care effectiveness indicators comprise statistics based on prevalence and incidence of drug users and prevalence and incidence of drug users being assisted by the health-care services. This may include statistics based on different critical periods in a drug user's career, such as the latency period from first heroin use to first treatment [1, 2] or the inter-relapse period [3].

However, since drug abuse is generally a stigmatized and hidden behaviour and is prosecuted in most countries, there have been varying degrees of underreporting when standard epidemiological survey techniques (such as the National Household Survey on Drug Abuse in the United States of America) have been used. It is thus necessary to develop methods that allow a more accurate estimation of the extent of the phenomenon from observational secondary data on drug abuse that are available in various forms. Estimation techniques refer for the most part to models and methodologies to calculate the extent and dynamics of drug abuse in a community and/or at the regional or national level based on various observed phenomena (secondary data) and on information received from certain target populations [4, 5]. Secondary data can be defined as existing statistical and documentary information that is routinely collected and readily available, such as treatment figures, drug seizures, infectious disease indicators or numbers of drug-related deaths.

Epidemics of problematic drug use

There are similarities between the spread of drug use, in particular for addictive drugs such as heroin, and the spread of infectious diseases. Use of drugs is communicated, obviously not as an organic agent, but as a kind of "innovative" social practice or custom, and not to everyone but only to those who, for whatever reason, are not immune (susceptible individuals). Once the basically contagious nature of drug use is accepted, it becomes possible to study and to model the process of transmission [3, 6, 7]. The epidemiological concepts of incidence (the rate of new cases occurring within a certain time period) and prevalence (the number of all existing cases at a certain moment in time) are thus operationally valuable

in studying illegal drug use. Unfortunately, as already mentioned, the population of users cannot be properly studied by standard statistical (descriptive) methods, since such users, because of drug laws and policies currently in force in various countries, are engaged in illegal acts. They therefore resist identification and constitute a "hidden" population. As a result it is necessary to use inferential methods and mathematical models in studying problematic drug use. Using indirect indicators such as numbers going for therapy, such models make it possible to estimate interesting quantities such as prevalence and incidence. Epidemiologists have suggested first use of drugs, first continuous use and first addiction as incidence indicators. Of these, the incidence of first use is the best measure of spread, even if there are ambiguities and difficulties in determining it, since not all cases are known and the definition of a "case" is itself problematic, possibly ranging from occasional use to addictive daily use [8]. Nevertheless, incidence of first use retains special significance as the most suitable measure of the tendency of new individuals to become involved with addictive drugs. Incidence analysis can thus be used as a tool for examining the process of diffusion of drug use. In particular, incidence figures may provide an indication of whether the number of problematic drug users is growing (epidemic phase), falling or stable (endemic phase).

The authors' main objective is to establish an appropriate model in order to develop an estimation method to study the behaviour of such epidemiological indicators. The extended application of the method reported below has revealed its major characteristics and usefulness in relation to policy issues.

Recently developed techniques of examining the epidemiology of infectious diseases that utilize surveillance data to estimate incidence and prevalence are now available and can be adapted in order to obtain more accurate results and estimates. Certain Back Calculation methods developed seem particularly useful.

Back Calculation has evolved considerably in relation to work on acquired immunodeficiency syndrome (AIDS) projections. Briefly, knowledge of the numbers infected with the human immunodeficiency virus (HIV) and the incubation period distribution are used to predict the number of AIDS cases that are expected in the future (direct or forecast use); on the other hand, if the number of AIDS cases is known and information on the incubation period distribution is available, estimates for those previously infected with HIV can be obtained (indirect or inferential use). Since the proportion of those infected who will eventually develop AIDS is unknown and as the Back Calculation is based on numbers of diagnosed AIDS cases, the method provides estimates only for those infected who will eventually develop the disease. That number, however, is what health authorities need to base their financial planning on. The present article shows how the same method can be utilized to back calculate historical trends in incidence of problematic drug use, more specifically in incidence of first use, from first reports of drug users in treatment. Based on treatment data, the study described here provides accurate estimates of the incidence of drug users who will eventually enter treatment. This is clearly a highly selected sub-population, since it excludes those individuals who will stop using drugs or die before starting any treatment, or those who will never develop problems requiring treatment. However, the

estimated incidence, which is an estimate of the number of those who will require treatment in the near future (the median lapse of time between the start of drug use and the first treatment has been estimated at five to six years), is appropriate for analysing spatial and time trends of problematic drug use, for planning, monitoring and evaluating preventive interventions and for forecasting treatment needs.

Estimating the incidence of problematic drug use: the Back Calculation approach

Back Calculation is a general type of deconvolution method originally proposed as a tool for estimating the minimum number of HIV-infected people and making short-term projections of AIDS incidence [9]. The basic idea of Back Calculation is to use a compilation of AIDS incidence data, an estimate of the incubation period distribution, usually externally obtained from a target population, in order to gain information about past HIV infection rates, which are usually unknown.

It is possible to use the same approach to back calculate, from observed data on drug users going for treatment, the incidence of onset of problematic drug use (same as for HIV infection) for those who will eventually be observed. If, for example, information is available about users first attending treatment (observed incidence), and if the actual incubation period or latency period distribution, that is, the time lapse between first use and first treatment are known or can be estimated, then it is possible to back calculate the incidence of first use on the basis of the known numbers of observed drug users. In other words, the Back Calculation makes it possible to estimate the incidence of drug users eventually seeking treatment using data on the incidence of drug users in treatment and an estimate of the latency period distribution, obtained externally from a suitable sample (target population).

This method has been applied in the study of the HIV/AIDS epidemic using various explicit formulations, corresponding to specific assumptions as to the expected incidence of AIDS, the shape of the HIV infection curve, the incubation period distribution and the estimation procedure. Each different combination of the above assumptions results in a different method of Back Calculation.

Heisterkamp and others [10, 11] proposed a Back Calculation method based on an Empirical Bayesian approach that had recently been adapted for estimating the incidence of first use of drugs on the basis of therapy incidence data [12, 13].

The therapy incidence data and the data needed to estimate the latency period distribution should be provided by the health-care services offering treatment (of any kind) to drug users. In particular, the present version of Empirical Bayesian Back Calculation uses aggregated biannual incidence data of "new" individuals under treatment in health-care services (at their first treatment). These are classified according to date of first registration with the health-care service and any other variable that could be used as a covariate or stratification variable, such as gender, age, geographical area, educational level and type of health-care service.

Similar individual (non-aggregated) data can be used to obtain an estimate of the latency period distribution. This is not only needed for applying Empirical

Bayesian Back Calculation, but it is also a simple tool to monitor the attractiveness of health-care services and harm-reduction programmes [1, 2].

A case study: the heroin epidemic in Italy

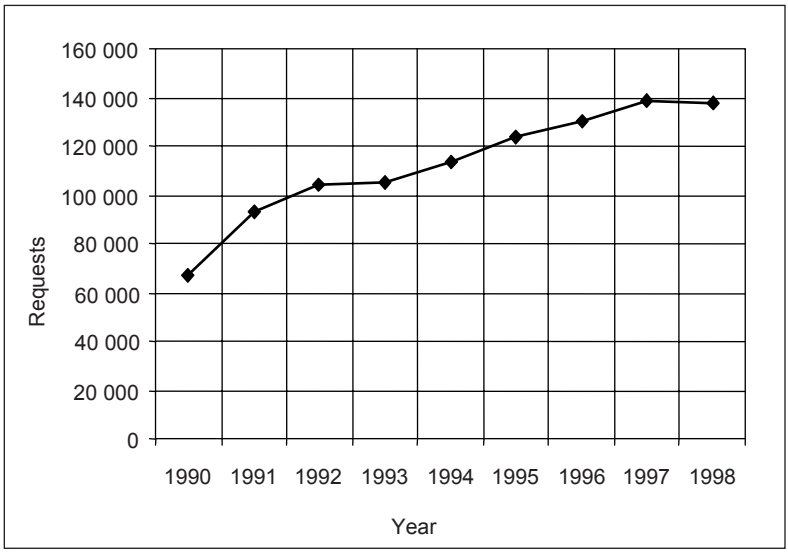
During the 1980s and 1990s, heroin by injecting caused the majority of interventions by both the health and the criminal justice departments in Italy [14-16]. The number of people in Italy between the ages of 15 and 54 who have used heroin at some time in their life is estimated to be not less than 300,000 [17, 18]. A description of the heroin epidemic based on official statistics follows, with the Empirical Bayesian Back Calculation method being applied in order to estimate the onset incidence curve of heroin use.

Observable data analysis

The estimated number of heroin users is almost double the number of people in treatment within the public services (based on returns from 518 public treatment services, representing 95 per cent of those services nationally and 100 per cent of their clients). The majority of treatment requests (about 90 per cent, on average, in the last 10 years) relate to heroin as the primary drug abused, since the beginning of public service activities (mid-1980s). Figures I-VII report the time trends of some interesting observable indirect indicators of problematic drug use over the last decade [16].

In terms of age of those in treatment, the trend is broadly towards an ageing population of drug users, as shown in figures I-VII (see also figures VIII-XIII. This

Figure I. Requests for treatment in Italy, 1990-1998



may be due to a stabilizing or decreasing trend of the onset incidence of heroin use in the last decade at the national level. This hypothesis is contradicted in part by the data from the Ministry of the Interior related to people referred for unlawful possession of drugs (any), reported below, even though the young age of people involved in these latter events may also be explained by the fact that most of them relate to possession of cannabis.

Figure II. All treatments in Italy, by age group, 1994-1998

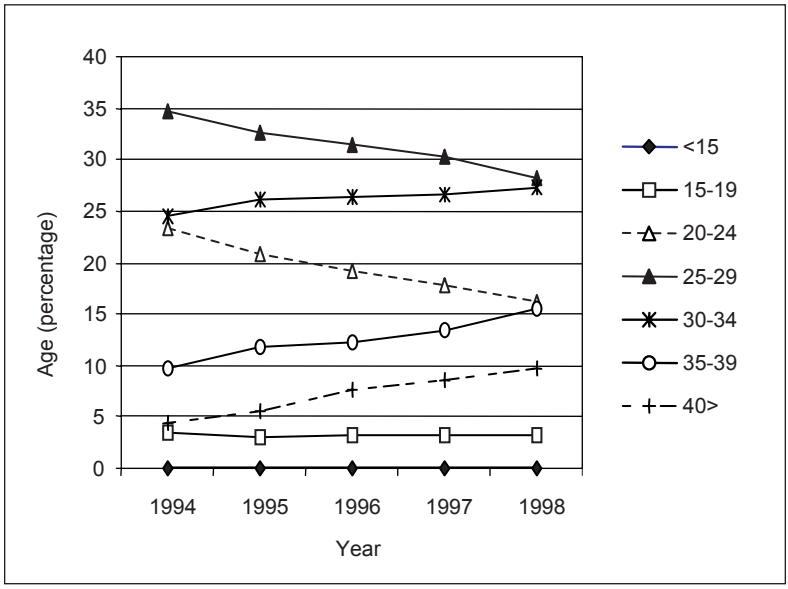


Figure III. First treatments in Italy, by age group, 1994-1998

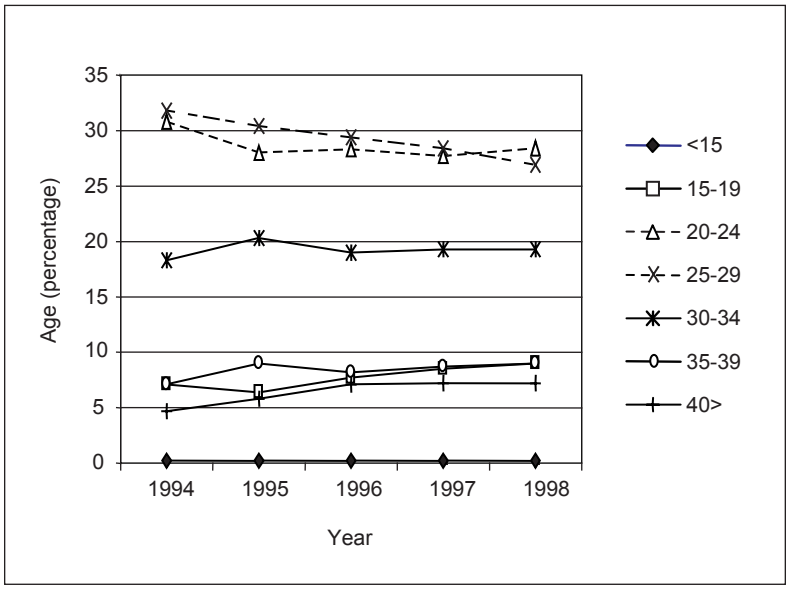


Figure IV. All requests for treatment in Italy, by age group, 1990-1998

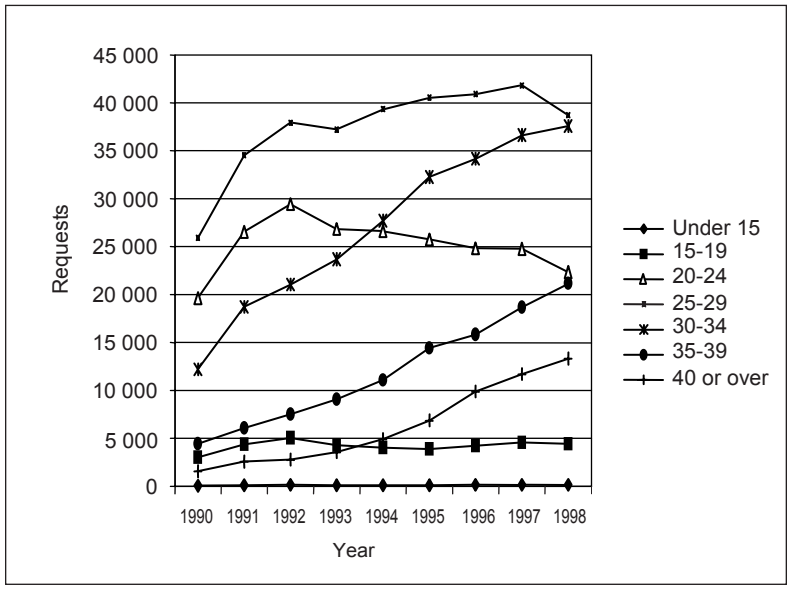


Figure V. First request for treatment in Italy, by age group, 1990-1998

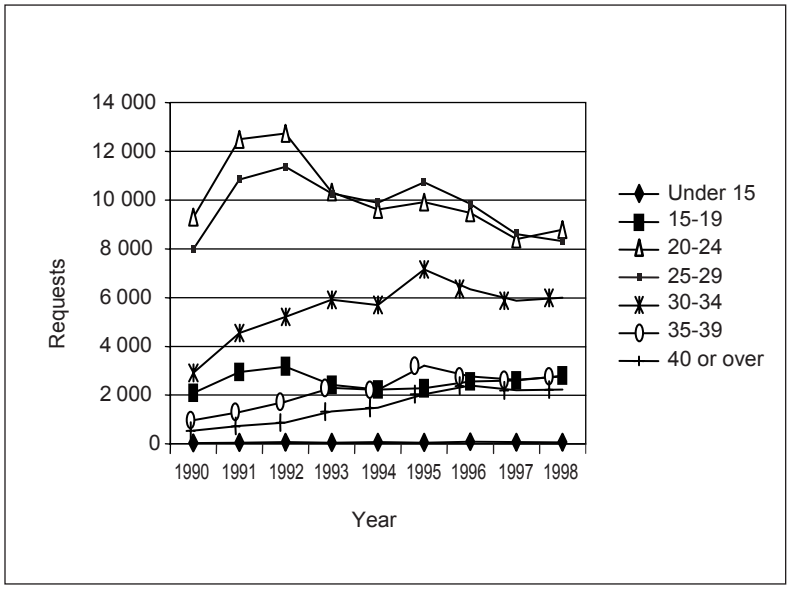


Figure VI. Males referred for unlawful possession of drugs in Italy, by age group, 1997 and 1998

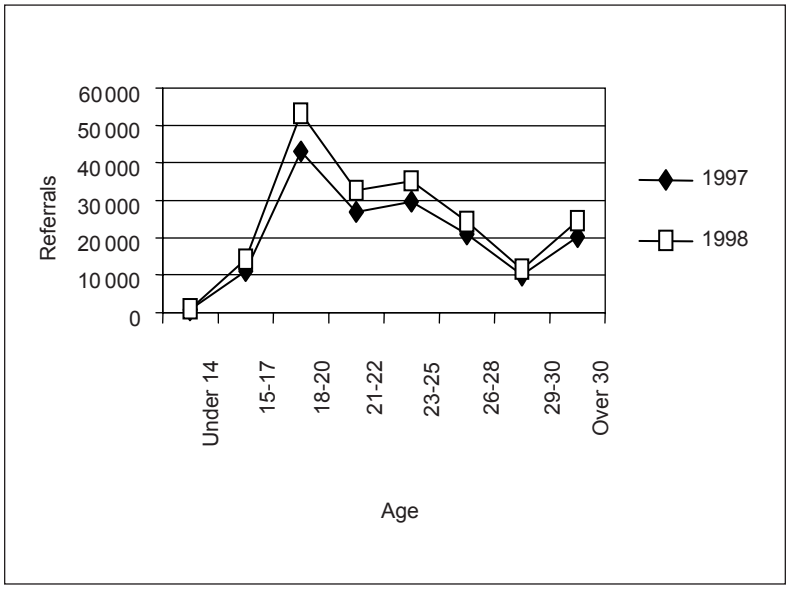
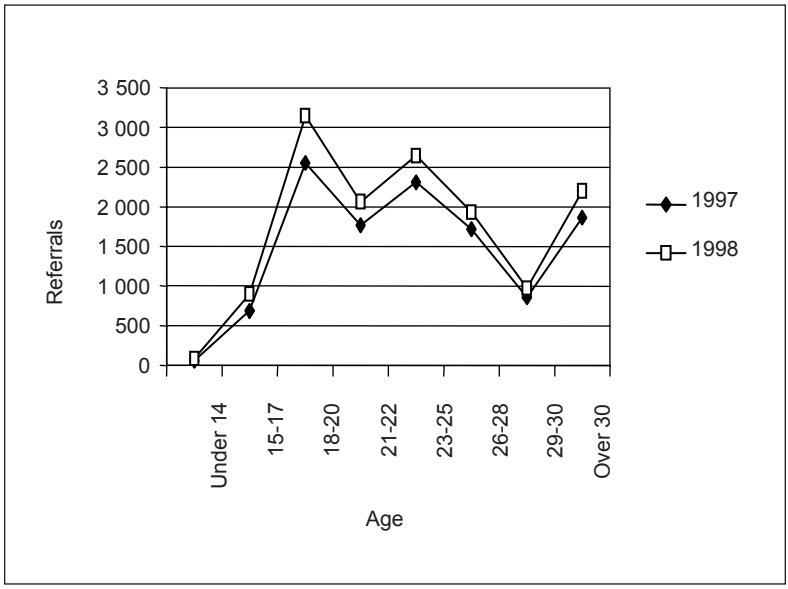


Figure VII. Females referred for unlawful possession of drugs in Italy, by age group, 1997 and 1998



In terms of drugs possessed unlawfully, cannabis was by far the most common, representing 61 per cent of all possession in 1997 and 64 per cent in 1998. A considerable way behind came heroin, followed by cocaine, Ecstasy and other opiates. Between 1997 and 1998, there was an increase in unlawful possession of all controlled drugs. The most significant rises were for possession of cannabis, cocaine, amphetamines and other drugs (primarily benzodiazepines). There were also important but smaller increases in possession of methadone and Ecstasy. Table 1 shows the distribution by drug, sex and year of people referred for unlawful possession. It should be noted that table 1 refers to the drugs that a person possessed and is therefore a greater number than the number of persons found in unlawful possession.

More than 50 per cent of heroin users in treatment are concentrated in six regions, which, in general, are the most heavily populated (see table 2). However, there are variations between regions related to the relative impact both of problematic drug use in general and of heroin use in particular (figures VIII-XIII). The pattern of primary drug use is broadly consistent throughout the country, with heroin the dominant drug of misuse among drug users in treatment. Variations between regions can also be observed regarding this aspect (table 2). It is not clear why these differences exist between regions. They may reflect relative availability of drugs in different regions, law enforcement activities, treatment policy and practice or different attitudes towards drug use in the regions, as well as differences in the socio-economic situations. It is also possible that the dynamics of the heroin epidemic in the different regions reflect different behaviours resulting from the pressure of the black market and to the influence of the trafficking routes. The estimation of the onset incidence, by means of the Back Calculation method described above, both at the national and at the regional level, provides a better understanding of those differences and of the data, as shown below. In particular, a higher incidence of treatments with respect to the resident population is to be expected in regions with an older epidemic (table 2, fifth column). In other words, possible peaks in the estimated incidence curves are anticipated in those regions with higher rates of new treatments per 1,000 residents (table 2, fifth column).

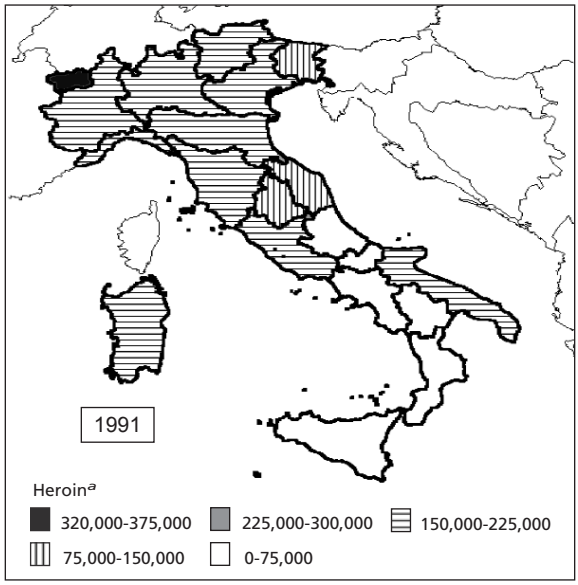
Table 1. Distribution by drug, sex and year of people referred for unlawful possession

	1997			*1998*		
Drug	*Male*	*Female*	*Total*	*Male*	*Female*	*Total*
Heroin	59 248	7 127	66 375	63 480	7 554	71 034
Methadone	579	113	692	676	122	798
Morphine	264	40	304	289	43	332
Opiates	2 317	292	2 609	2 329	292	2 621
Cocaine	9 516	684	10 200	11 750	849	12 599
Amphetamine	438	27	465	562	35	597
Lysergic acid diethylamide (LSD)	420	41	461	468	46	514
Ecstasy and analogues	2 748	221	2 969	3 192	250	3 442
Cannabis	126 720	6 364	133 084	156 254	7 920	164 174
Other drugs	765	82	847	1 223	120	1 343

Table 2. Distribution by drug users in treatment in Italy, by region, 1989

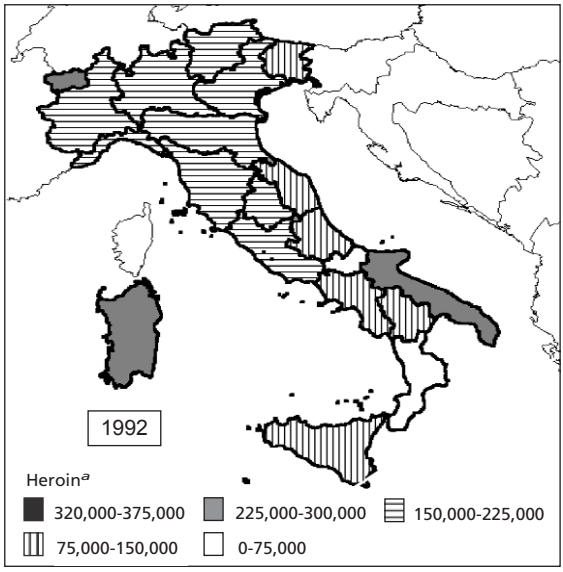
Region	*Residents (millions)*	*Drug users in treatment*	*Heroin as primary drug (percentage)*	*Treatments per 1 000 residents*
Piedmont	4.3	13 083	92.9	3.04
Valle d'Aosta	0.1	338	97.3	3.38
Lombardy	9.0	20 868	86.8	2.32
Trentino-Alto Adige	0.9	1 558	92.0	1.73
Veneto	4.5	11 032	79.6	2.45
Friuli-Venezia Giulia	1.2	2 591	77.9	2.16
Liguria	1.6	5 546	83.8	3.47
Emilia Romagna	4.0	8 942	85.8	2.24
Tuscany	3.5	9 673	86.5	2.76
Umbria	0.8	1 897	88.7	2.37
Marches	1.5	4 110	76.0	2.74
Latium	5.2	11 013	91.0	2.12
Abruzzo	1.3	3 171	75.2	2.44
Molise	0.3	534	80.5	1.78
Campania	5.8	11 327	81.2	1.96
Apulia	4.1	13 178	80.3	3.21
Basilicata	0.6	893	91.4	1.49
Calabria	2.1	4 294	81.1	2.04
Sicily	5.1	8 103	89.4	1.59
Sardinia	1.7	5 506	96.9	3.24
Total, Italy	57.6	137 657	88.5	2.38

Figure VIII. Drug users in treatment in Italy, by region, with heroin as the drug of primary use, 1991



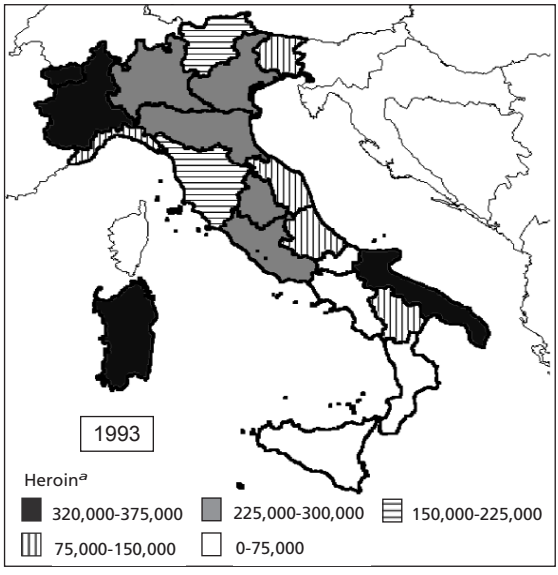
[a]The values give the number of treatments per 100 million residents (e.g. 100,000 means 0.1 per cent)

Figure IX. Drug users in treatment in Italy, by region, with heroin as the drug of primary use, 1992



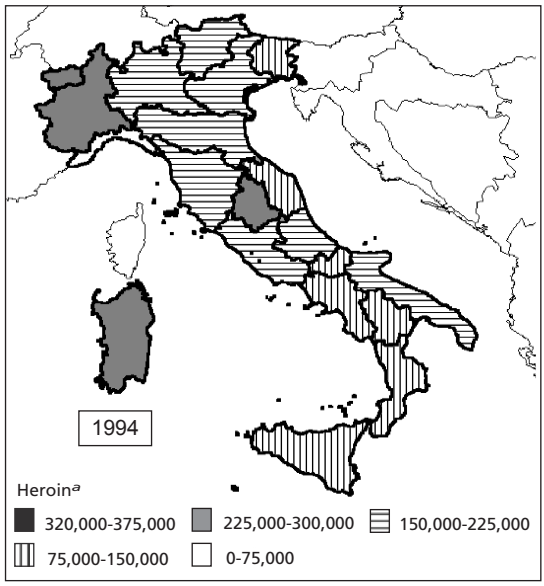
[a]The values give the number of treatments per 100 million residents (e.g. 100,000 means 0.1 per cent)

Figure X. Drug users in treatment in Italy, by region, with heroin as the drug of primary use, 1993



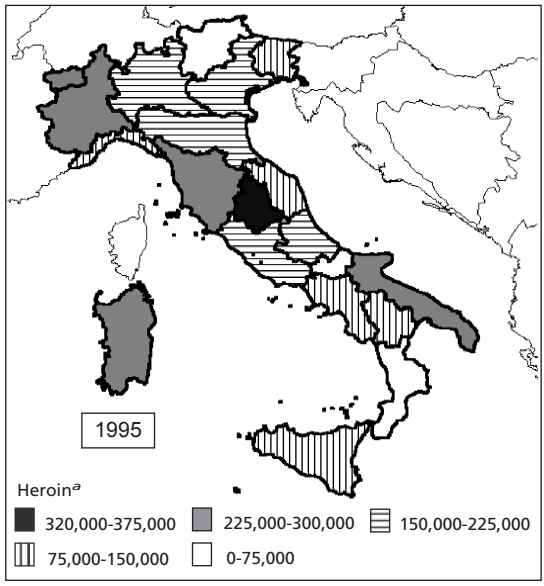
[a]The values give the number of treatments per 100 million residents (e.g. 100,000 means 0.1 per cent)

Figure XI. Drug users in treatment in Italy, by region, with heroin as the drug of primary use, 1994



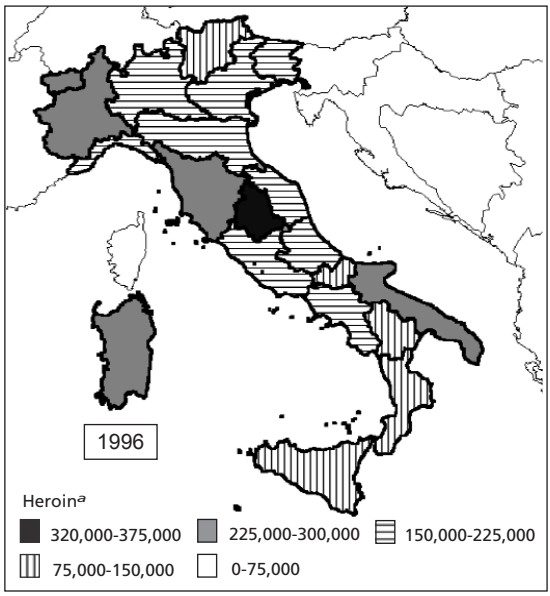
[a]The values give the number of treatments per 100 million residents (e.g. 100,000 means 0.1 per cent)

Figure XII. Drug users in treatment in Italy, by region, with heroin as the drug of primary use, 1995



[a]The values give the number of treatments per 100 million residents (e.g. 100,000 means 0.1 per cent)

Figure XIII. Drug users in treatment in Italy, by region, with heroin as the drug of primary use, 1996



[a]The values give the number of treatments per 100 million residents (e.g. 100,000 means 0.1 per cent)

Estimating the hidden onset incidence from observed therapy incidence

Empirical Bayesian Back Calculation was first applied to national data provided for the period 1986-1998 by the Ministry of Health, which routinely collects data on clients in public and private services. Data on clients in private services were excluded as those clients are mostly referred from public centres and double counts would have been high. Data include only clients who checked in for treatment for the first time and do not include those who only contacted services but did not receive treatment (see figures I-VII). Treatment refers to any therapeutic and rehabilitation procedure, whether pharmacological or not, offered by the service, even outside the premises (prisons, therapeutic communities or hospitals). Public treatment centres are thought to provide about 50 per cent methadone-substitution treatments and 50 per cent psychiatric and social interventions, with high interregional variability [19].

The data were classified according to age group, gender and region. The latency period was estimated using various local data sets provided by the different regional health authorities containing individual data on drug users in treatment. From the results obtained, the period appears to be remarkably similarly

distributed over the different sites, with a median of between five and six years and an average of between five and seven years. Figure XIV reports the survival curves, estimated by the Kaplan-Meyer method [1, 12], only for the data related to the five provinces of the Latium region. That time lapse, however, appears to be much longer than in young drug users, suggesting that age at first use is an important covariate to be included in any model [2, 14]. For each data set, the best parametric estimate of the latency period distribution was obtained by using the P-P plot method (quantile best fitting), which is available in the Statistical Package for the Social Sciences (SPSS) used for the present application. The best model resulted, for each site considered, in either a Gamma or a Weibull density (figures XVI and XVII) with slightly different parameters [12]. On the basis of a sensitivity analysis [13], based also on further secondary data, the most representative estimates at the national level were achieved in a Gamma density, with parameters 1.51 (shape) and 0.30 (scale), and a Weibull density, with parameter 1.28 (shape) and 5.33 (scale) (figure XV). The Gamma, being less dispersed than the Weibull, is always associated with smoother incidence curves resulting from the estimation procedure based on Empirical Bayesian Back Calculation. Further results and discussion of the latency period analysis are reported elsewhere [2].

In order to apply Empirical Bayesian Back Calculation, the original therapy incidence data were first multiplied by 0.70 to take into account the estimated proportion of double counting and the proportion of drug users in treatment who are not heroin users (about 10 per cent).

Figure XIV. The latency period: Kaplan-Meyer survival functions

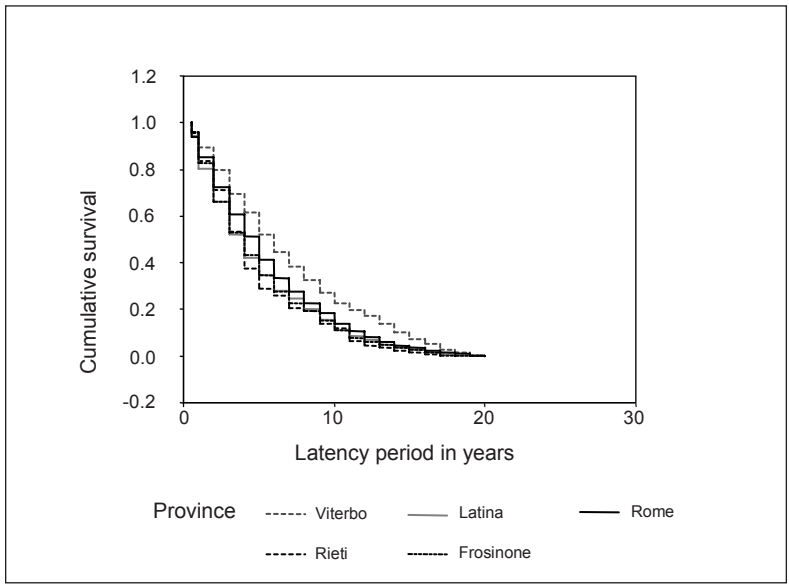


Figure XV. The latency period: Gamma and Weibull distributions

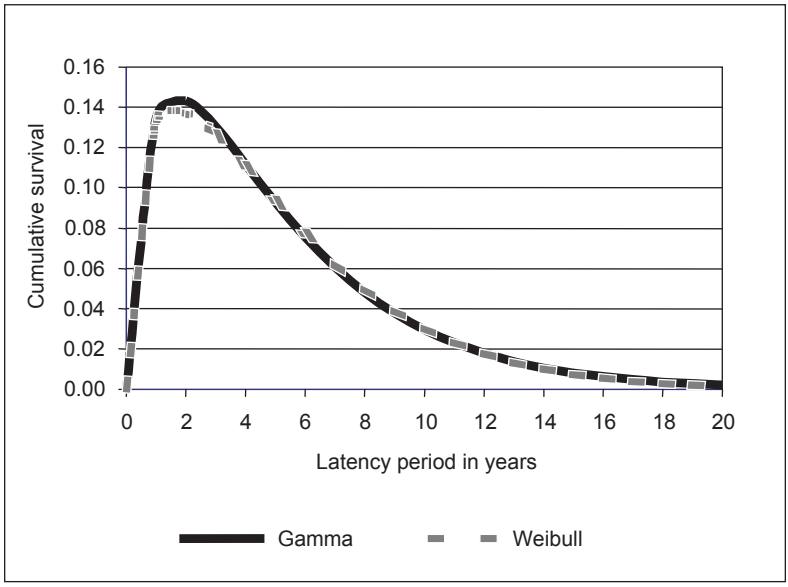


Figure XVI. Gamma P-P plot of the latency period

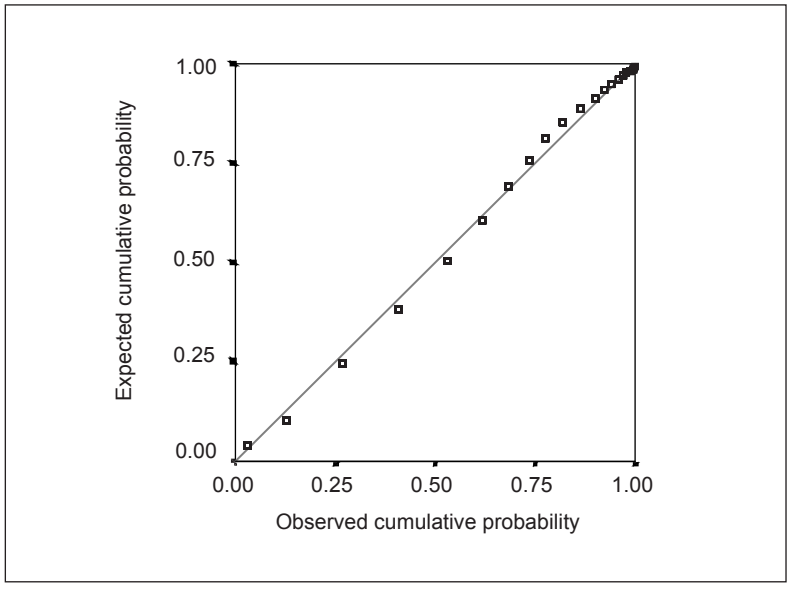
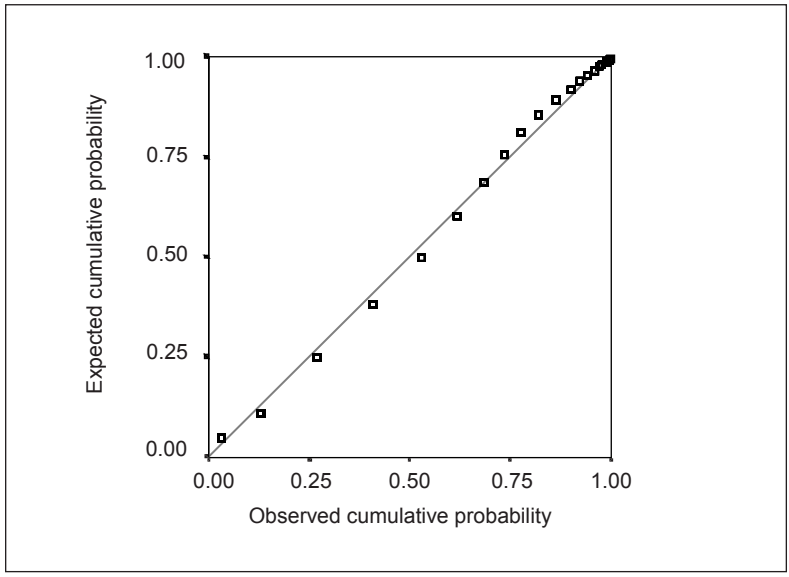


Figure XVII. Weibull P-P plot of the latency period



The results of the estimation are shown in figures XVIII and XIX. Figure XVIII reports the incidence curve of drug users estimated by Gamma and the Weibull latency period distribution, with and without age at first treatment as a covariate. Figure XIX reports the incidence curve of drug users under treatment estimated by Gamma and Weibull latency period distribution, with and without age at first treatment as a covariate; the observed therapy incidence data are represented by dots.

The incidence curves provided by the Empirical Bayesian Back Calculation estimation procedure are dependent on the latency period model chosen, but the location of the peaks of the epidemic seems to be a reliable parameter. The qualitative trends also seem to be reliably estimated, as shown below.

From information about heroin prevalence among males aged 18 provided by the annual surveys of military conscripts, information about distribution of age at the start of heroin use and overall prevalence estimates for recent years [20], it is possible to predict that the incidence curve will not show high oscillations. Figure XVIII indicates that, when the model is applied without the age covariate, the estimated curves show a higher oscillation than when the covariate is included. This suggests that the age covariate should always be included in the model, confirming the fact that age is a factor that strongly influences the latency period [2, 14]. Moreover, though the Gamma and Weibull distributions result in similar incidence curves, the Gamma curve is slightly smoother than the Weibull, as expected. Thus, the model based on the Gamma distribution with the age covariate included seems to provide the best representation of the epidemic. The incidence curve shows two peaks, the lowest peak in 1985 and the highest one in 1991. Nevertheless, this is not necessarily evidence of the existence of two different epidemics, since the two peaks could be determined by the combination of different local sub-epidemics developing differently over time.

Figure XVIII. Incidence of drug users in the total population of Italy, estimated by Gamma and Weibull latency period distribution, with and without age at first treatment as a covariate, 1975-1998

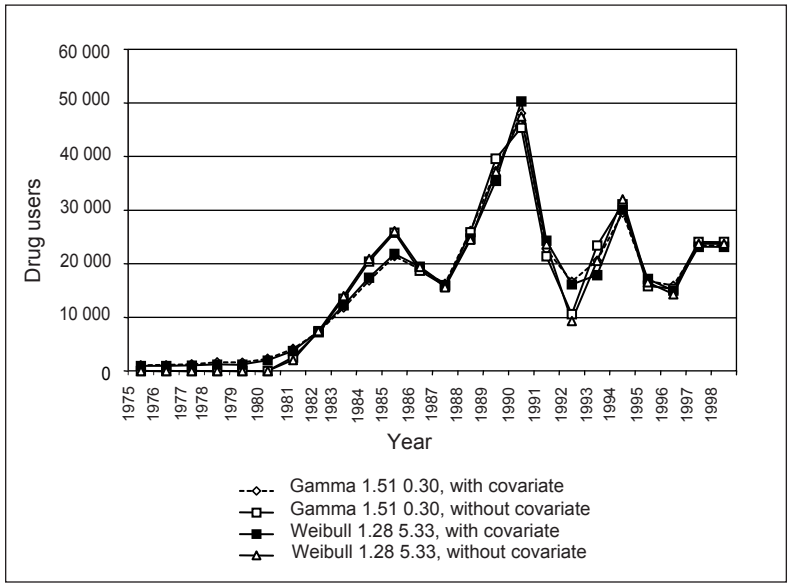
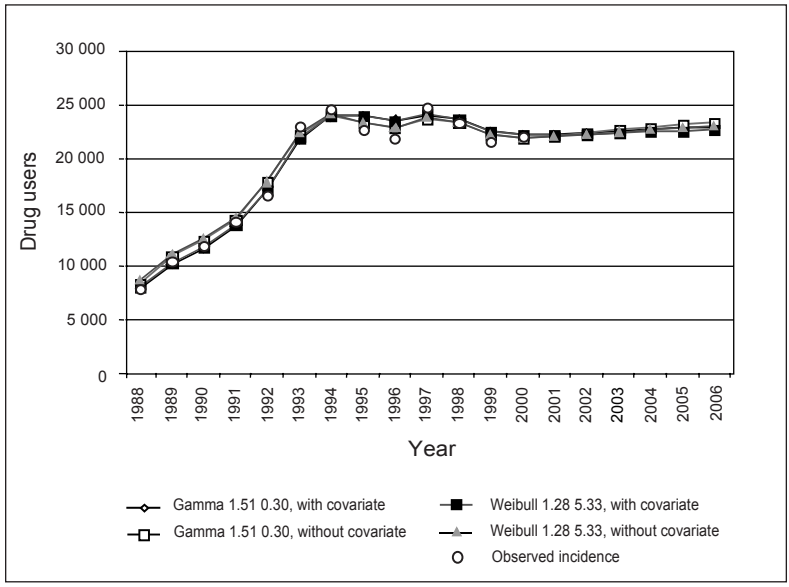


Figure XIX. Incidence of drug users in treatment in Italy, estimated by Gamma and Weibull latency period distribution, with and without age at first treatment as a covariate, total population, 1988-2006



The same trend, in particular the location of the peaks, can also be observed for the incidence curve estimated for males and females separately [19]. Nevertheless, the general trends for the two sexes are quite different: in particular, the second peak is higher for males and the first lower for females. This is in agreement with the observed trend in the male/female ratio in the therapy data available. The estimated overall size of the epidemic among the population under study, measured by the cumulative incidence from 1975 to 1998, is about 385,000. It is important to stress that, as mentioned above, that number underestimates the total size of the epidemic during that period, owing to the hidden population of drug users who will never request therapy. The number is in agreement with the perception of the phenomenon found in other sources and studies reporting prevalences in various years [18-20].

It must be stressed again that it is important to improve the information given by the Back Calculation estimate of recent incidence using all the external data available from different sources, for example, the information provided by the surveys of military conscripts made it possible to exclude improbable results obtained when using particular latency distribution models.

It is important to observe that, after a period of rapid increase during the 1980s, the incidence curve seems to stabilize at the beginning of the 1990s. That trend may indicate the success of the prevention interventions, but could also reflect the saturation effect of the population at risk expected in all epidemic processes [3]. The effect could also be due to the fact that the national epidemic is a combination of several local epidemics, in which case the national estimate of the incidence curve would simply represent the sum of the local estimates. As a matter of fact, from further analyses reported below (figures XX-XXII), it appears that the application of the method at the regional level provides a picture of several sub-epidemics with peaks located at different times and with clear space-time trends. Figure XX shows the onset incidence estimates for six of the largest regions in Italy. As can be seen, the principal peak of each curve varies considerably. In particular, the northern regions (Lombardy and Emilia Romagna) peaked before the others, more or less simultaneously the peak of Apulia can be observed (this region can be considered a border region with respect to the Balkan trafficking route), then the peaks of the central and central-southern regions (Latium and Campania) occur and, finally, the peak in Sicily, where, owing to the wider uncertainties corresponding to recent years, it is also possible that the curve is still increasing. Statistics on deaths from drug abuse confirm this last observation for Sicily. Figure XXI gives forecasts of therapy incidence in the same regions, which can be used to estimate health-care needs in the short and medium term. The observed therapy incidence data correspond well to the curves shown in the graph (not shown in the present paper). Figure XXII reports the estimated cumulative incidences, which measure the overall impact of the epidemic since the beginning in each region. Further analyses for Italy are reported elsewhere [19}.

The authors' findings show that, in the northern and border regions, where the heroin epidemic is older, the most cost-effective interventions that should be planned are related to health care and rehabilitation, whereas in the regions where the epidemic started later, in Sicily, for example, prevention efforts can also still have a large impact. Further comments on policy implications are reported

elsewhere [19]. The Empirical Bayesian Back Calculation method has recently been applied to therapy incidence data from Amsterdam and quite interesting results and comparisons have been obtained [13].

Figure XX. Incidence of drug users in six regions in Italy, estimated by Gamma latency period distribution, with age at first treatment as a covariate, total population, 1975-1998

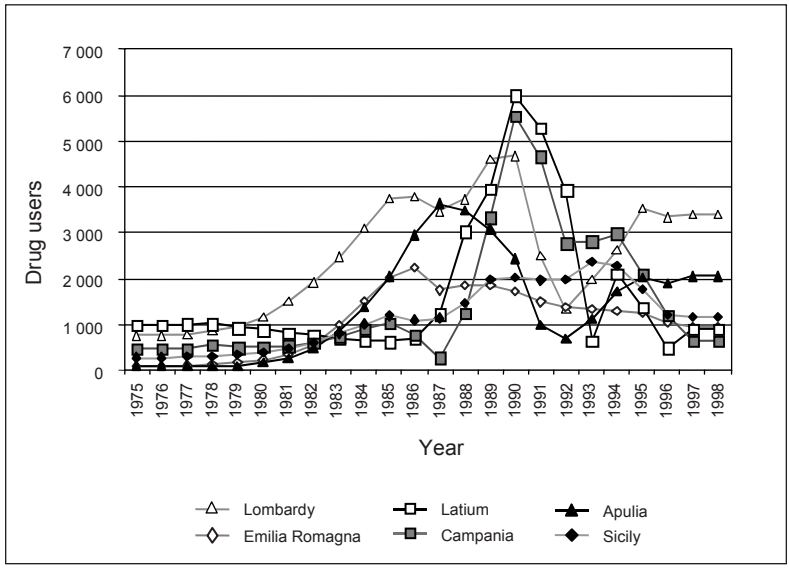


Figure XXI. Incidence of drug users in treatment in six regions in Italy, estimated by the Gamma latency period distribution, with age at first treatment as a covariate, total population, 1988-2006

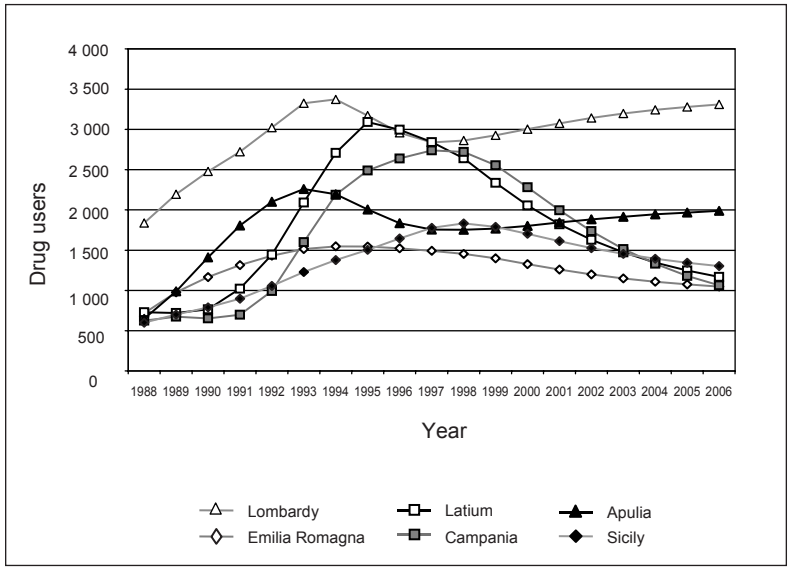
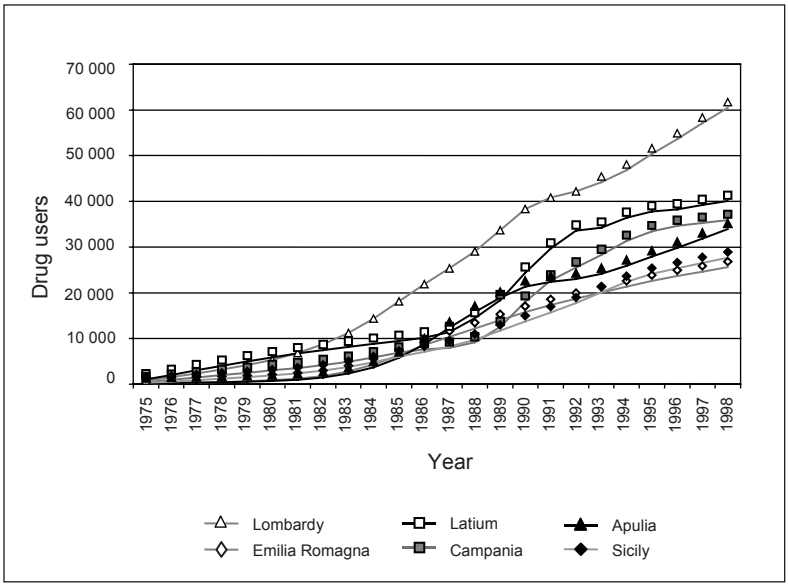


Figure XXII. Cumulative incidence of drug users in six regions in Italy, estimated by the Gamma latency period distribution, with age at first treatment as a covariate, total population, 1975-1998



References

1. C. Rossi, "Monitoring drug control strategies: hidden phenomena, observable events, observable times", *International Journal of Drug Policy,* vol. 10, No. 1 (1999), pp. 131-144.
2. L. Wiessing and others, "Analysis of the latency period from onset of heroin use to first treatment: comparisons among four European sites, possible implications for treatment services and for understanding the natural history of problematic heroin use", forthcoming.
3. C. Rossi, "A mover-stayer type model for epidemics of problematic drug use", *Bulletin on Narcotics* (United Nations publication), vol. LII, Nos. 1 and 2 (2001).
4. United Nations International Drug Control Programme, *Drug Abuse Rapid Situation Assessment and Responses* (United Nations, Office for Drug Control and Crime Prevention, Vienna, 1999).
5. C. Rossi, "Using available data with dynamic models to estimate the extent of problem drug use", in *Modelling Drug Use: Methods to Quantify and Understand Hidden Processes,* European Monitoring Centre for Drugs and Drug Addiction Scientific Monograph Series, no. 6, 2001, pp. 67-90.
6. D. A. Behrens and others, "A dynamic model of drug initiation: implications for treatment and drug control", *Mathematical Biosciences,* vol. 159, No. 1 (1999), pp. 1-20.
7. L. Billard and P.W.A. Dayananda, "Drug addiction-pushers generated from addicts", *Biometrical Journal,* vol. 35, 1993, pp. 227-244.

8. L. G. Hunt and C. D. Chambers, *The Heroin Epidemics* (New York, Spectrum Publications, 1976).
9. R. Brookmeyer and H. G. Gail, "Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States", *The Lancet,* vol. 2, 1986, pp. 1320-1322.
10. S. H. Heisterkamp, A. M. Downs and J. C. van Houweling, "Empirical Bayesian estimators for reconstruction of HIV incidence and prevalence and forecasting of AIDS: I. Method of Estimation", in *Quantitative Analysis of HIV/AIDS: Development of Methods to Support Policy Making for Infectious Disease Control,* Ph.D. thesis, University of Leiden, 1995, pp. 65-98.
11. S. H. Heisterkamp, J. C. van Houwelingen and A. M. Downs, "Empirical Bayesian estimators for a Poisson process propagated in time", *Biometrical Journal,* vol. 41, No. 1 (1999), pp. 358-400.
12. European Monitoring Centre for Drugs and Drug Addiction, "Pilot project to estimate time trends and incidence of problem drug use in the European Union", final report, Lisbon, 1999.
13. L. Ravà and others, "Estimating the incidence of problem drug use by back calculation methods", forthcoming.
14. European Monitoring Centre for Drugs and Drug Addiction, *1999 Annual Report on the State of the Drugs Problem in the European Union,* Lisbon, 1999.
15. United Nations International Drug Control Programme, *World Drug Report* (Oxford, Oxford University Press, 1997).
16. Italian Focal Point, *Annual Report on the State of the Drugs Problem in Italy for the Year 1999* (Rome, 2000).
17. L. Ravà and C. Rossi, "Estimating the size of a hidden population involved in the HIV/AIDS epidemic: a method based on back-calculation and dynamical models", in *Simulation in the Medical Sciences,* J. G. Anderson and M. Katzper, eds. (San Diego, California, Society for Computer Simulation International, pp. 57-62, 1999).
18. C. Rossi, "Estimating the prevalence of injecting drug users on the basis of Markov models of the HIV/AIDS epidemic: applications to Italian data", *Health Care Management Science,* vol. 2, 1999, pp. 173-179.
19. L. Ravà and others, "The heroin epidemic in Italy: space-time modelling, data analysis and forecasts", forthcoming.
20. Presidenza del Consiglio dei Ministri, *Relazione Annuale al Parlamento sullo Stato delle Tossicodipendenze in Italia 1999* (Annual report to Parliament on the state of drug addictions in Italy 1999) (Roma, 2000).