Modelling techniques and their uses

INTRODUCTION

In Chapter 6, Comiskey and Hay present the back-calculation (BC) method as a tool to estimate unobserved incidence of a disease from observed data and to account for a latency time to onset of disease. This method has been extensively used and further developed to estimate HIV incidence from observed AIDS data. The authors discuss the application of BC models to drugs data, for example, for estimating onset incidence of problem use from observed treatment data and the need for research into the 'incubation time' or latency time from first drug use to first treatment.

Hay and Comiskey discuss in Chapter 7 the application of compartmental models in different fields, notably infectious-diseases epidemiology, and their usefulness for estimating drug use. They introduce the concept of deterministic and stochastic models with their different implications for results (existence or not of random variation) and method of analysis (mathematical versus computational). The authors raise relevant questions about the applicability of models to describe the spread of drug use, given the basic uncertainty and lack of consensus about underlying mechanisms and definitions and lack of adequate data.

In Chapter 8, Taylor and Young distinguish static and dynamic models, defining the latter as those in which a population is traced over time. A series of basic issues concerning dynamic modelling are discussed, including time frame and geographic boundaries of a prevalence-estimation exercise, types and levels of drugs used and the relationship between definitions of drug use and the use of data sources. The authors also describe and evaluate a number of modelling techniques adopted in this class of model. Finally, system-dynamic modelling is discussed, which describes an interconnected set of endogenous feedback mechanisms (time movements and accumulations of people, materials and information), most of which are not observable but which generate observable output over time.

In their following chapter, Taylor and Young describe the use of factor analysis to either confirm or explore underlying latent variables from observed variables. The authors argue that working with composite latent variables is often nearer to reality than interpreting use of each drug separately, as most drug users are polydrug users and there are different and combined ways of using (e.g. moderate or heavy alcohol use in combination with other drugs). Taylor and Young then discuss structural-equation modelling and conclude that these highly complex models are difficult to manage and are of less practical value, but are a potentially very useful explorative tool to understand how underlying theoretical concepts can be fitted into models of drug prevalence.

CHAPTER 6

THE METHOD OF BACK CALCULATION AS A MEANS OF ESTIMATING INCIDENCE

Catherine Comiskey and Gordon Hay

The method of back calculation uses knowledge of the numbers of AIDS cases (after adjustment for reporting delays) and information on the incubation period distribution to derive estimates of those previously infected with the HIV virus. This method has been widely adopted by statisticians and mathematicians working in the United States and in Europe as a means of providing essential estimates of the unknown numbers who are HIV infectious and short-term predictions of future numbers with AIDS. This chapter provides a detailed, non-technical description of the method. Working with the first estimates based on Irish AIDS cases, the anomalies in the method with respect to the Irish data are illustrated. They are then confirmed by comparing the results with those derived in the UK and US and highlighting the limitations inherent in the approach. It is discussed where the method has been applied in the past and suggestions are made for future model developments and applications within drug research. A critical assessment of the model type is given for future use as well as recommendations on its use and feasibility.

Model description

Knowledge of the numbers infected with the HIV virus and of the incubation period distribution allows the number of AIDS cases which can be expected in the future from these HIV patients to be predicted. Conversely, if the number of AIDS cases (after adjustment for reporting delays) is known and information on the incubation period distribution is available, estimates for those previously infected with the HIV virus may be derived. The numbers previously infected can then be projected forward to obtain short-term predictions for the number of AIDS cases. This method is known as back calculation. It must be stressed, however, that since the proportion of those infected who eventually go on to develop AIDS is unknown, and as the number of diagnosed AIDS cases is being used, this method provides estimates only for those infected who will eventually go on to develop the disease. It is these disease numbers, however, that health authorities and policy-makers have to finance and plan for.

This relationship between the incidence of AIDS cases, the incidence of HIV infections and the distribution of the incubation period is well documented and may be expressed in the form

$$a(t) = \int h(t - u)\, f(u)\, du$$

where $a(t)$ is the rate of new diagnoses of AIDS cases, $h(t)$ is the rate of acquiring HIV infection and $f(t)$ is the distribution of the incubation period. The method of back calculation is based on the premise that if any two of these three quantities are known then the third can be derived. With the current AIDS epidemic both $a(t)$, the rate of new diagnoses of AIDS cases, and $f(t)$, the distribution of the incubation period, can be observed and measured, hence $h(t)$, the unknown number of HIV cases, at time t can also be derived. The choice of distribution for both $a(t)$ and $f(t)$ is, however, very varied and depends upon the stage of the epidemic and the availability of data. Estimates of HIV incidence derived from the back-calculation method will vary depending on the distributions used. This is discussed further in later sections.

It is important to remember that the number of diagnosed AIDS cases are adjusted for reporting delays. Reporting delays in countries vary and arise when a department of health acts as the centre for the collation of national data on AIDS, with diagnosed cases being reported directly to an AIDS coordinator. A considerable time lag may occur between the time that a case is diagnosed and the time at which it is recorded in the national AIDS figures (the difference between these times being known as the reporting delay). It is at the time of diagnosis and not the time of reporting that the patient will need treatment and it is therefore necessary to plan for and work with the number of diagnosed cases.

Where has the model been applied?

This approach to estimating the incidence of HIV infection and projections for the number of AIDS cases was first proposed in the United States (Brookmeyer and Gail, 1986) and in the United Kingdom (Isham, 1988; Day and Gore, 1988). Further discussions on the method applied to the US epidemic appeared in the late 1980s (Brookmeyer and Damiano, 1989). Back calculation as a means of predicting the incidence of HIV infection in the UK was discussed again in a review of mathematical and statistical studies of HIV (Anderson, 1989). More recently, this method was adapted to provide further predictions of the AIDS epidemic in the UK (Day et al., 1989). For application to the sparse Irish data the method was adapted further (Comiskey and Ruskin, 1992).

Similarly, the method has been applied across Europe, for example to reported AIDS cases in Amsterdam (Hendricks et al., 1992). The authors, like Comiskey and Ruskin, found that the choice of the exact model greatly influenced the predictions. In Italy, the method is also implemented to estimate the incidence and prevalence of HIV (Mariotto and Cascioli, 1996). Authors found that the choice of one particular back-calculation model over another among three different models considered seemed to be a minor component of the uncertainty when compared to the uncertainty due to all other

sources. Variations in model choice are also addressed in the literature (Cooley et al., 1996). These authors combine information from 12 different studies in an effort to estimate the AIDS incubation distribution with greater precision than is possible from a single study. In their conclusions they stress that back-calculation approaches should allow the incubation distribution to take several forms to adequately represent HIV estimation uncertainty.

The method has also been applied to the prediction of the HIV/AIDS epidemic in the low-prevalence countries of Central and Eastern Europe (Downs et al., 1997). This study included the Asian republics of the former Soviet Union. The authors found that recent reports of rapidly increasing HIV infection rates suggest that back calculation may seriously underestimate the size of the epidemic in low-prevalence countries. The application of the method has not been confined to Europe and the United States, but has also been used in Australia (Becker and Chao, 1994) and in the city of Perth in Western Australia (Comiskey, 1996). Becker and Chao (1994) improve the results from the method by assuming that quarterly HIV incidences are not independent.

To summarise, the back-calculation method has been applied globally and under diverse conditions. The core idea of the method remains unchanged in that the back-calculation is from known AIDS incidence, through the incubation-period distribution in order to estimate the unknown or hidden pool of HIV carriers. The implementation of the method, however, varies considerably between authors and its technical details are continually being refined. A key factor in the implementation of the method is the choice of the incubation-period distribution. Estimates of HIV incidence and prevalence derived from the method will vary considerably depending on the mathematical or statistical nature of the distribution used by the researcher. In spite of these reservations, the method continues to be employed and receives considerable attention in the academic press.

The method used for the example illustrated here is based upon the knowledge of the distribution of AIDS cases and the distribution of the incubation period (Brookmeyer and Gail, 1986; Brookmeyer and Damiano, 1989; Anderson, 1989). The sensitivity of this method to parameter estimates has been discussed in the literature (Wilson et al., 1992) and many authors have studied ways of improving the method by using either age-specific incidence rates or more accurate estimates of the incubation-period distribution by using dates of first positive HIV test (Rosenburg, 1994; Dietz et al., 1994; Marschner, 1994). In the United Kingdom, work has been carried out on the study of the incubation-period distribution (Anderson and Medley, 1988). The authors fit a gamma distribution with a mean of 14.3 years with approximately 40 % of all those infected progressing to AIDS within ten years. In this example, this choice of incubation distribution is implemented. The method is illustrated with data on Irish AIDS cases from 1982 to 1992. In the light of comments on the necessity for validation of HIV predictions (Bailey, 1994) the results from the improved back-projection method compare favourably with HIV predictions arising from transmission models of HIV and AIDS in Ireland.

Adjusting the number of diagnosed AIDS cases for reporting delays

For every AIDS case reported in Ireland, data are supplied on the risk group (and other demographic features) of the patient and the date that he or she was diagnosed as falling within the Centres for Disease Control (CDC) definition of an AIDS case. At the time the model was applied, the CDC classified those with the HIV virus into four groups: Group I, seroconversion; Group II, asymptomatic; Group III, progressive generalised lymphadenopathy; and Group IV, patients with clinical manifestations of HIV infection designated by assignment to one or more sub-groups (A–E). Within Group IV, sub-group classification is independent of the presence or absence of lymphadenopath. Sub-group IV C1 defined an AIDS case. It consisted of patients with secondary infectious diseases defined as the diagnosis of an infectious disease associated with HIV infection and/or at least moderately indicative of a defect in cell-mediated immunity. Included are patients with symptomatic or invasive disease due to one of a number of specified secondary infectious diseases.

The problem of reporting delays in the number of AIDS cases is discussed (Brookmeyer and Damiano, 1989). These authors base their method on a conditional likelihood for estimating the reporting delay distribution. Their method has been adapted to adjust the reported numbers of Irish AIDS cases (Comiskey and Ruskin, 1992).

Methods

The integral equation model arising in back projection is a linear Volterra equation of the first kind with a difference kernel

$$a(t) = \int h(t-u)\, f(u)\, du \qquad (1)$$

where $a(t)$ is the rate of new diagnoses of AIDS cases (adjusted for reporting delays), $h(t)$ is the rate of acquiring HIV infection and $f(t)$ is the distribution of the incubation period. Although equation 1 may be viewed as the convolution of $h(t)$ and $f(t)$, it is difficult to find a solution by means of inverse Laplace transforms when $f(t)$ is given by a gamma distribution with parameter $\alpha \varepsilon \mathbb{R}$. However equation 1 can be changed into a generalised Abel integral equation by differentiation. The solution is then given in terms of an integral in the two known functions $a(t)$ and $f(t)$. The resulting integral can then be solved in terms of incomplete and complete gamma functions and error bounds for all solutions can be provided (Comiskey, 1992).

Three different mathematical forms for the growth in the annual incidence of AIDS cases, $a(t)$, are considered here. It has been shown that in Europe the numbers of new AIDS cases grew rapidly in the early years of the epidemic (Downs et al., 1987). This has since been followed by a period of slower growth with doubling times increasing. Three forms for $a(t)$ are considered to describe the rate of new diagnosis of AIDS cases:

$$a(t) = a_0 exp(a_1 t - a_2 t^2) \qquad (2)$$

$$a(t) = (b_0 + b_1 t) / exp(1-t) + b_2 \qquad (3)$$

$$a(t) = c_0 + c_1 t + c_2 t^2 \qquad (4)$$

These three forms reflect empirical results up to 1990 which suggest an initial period of exponential growth in AIDS cases. The first growth rate for *a(t)*, the quadratic exponential, describes rapidly increasing AIDS incidence from 1990. The second, the linear logistic, describes decreasing incidence from 1990 and the third, the quadratic, describes steadily increasing incidence from 1990. These are chosen not to model the future incidence of AIDS cases, but rather to model the range of values for which to estimate *h(t)*.

The solution of 1 is considered when

$$f(t) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\lambda t) \quad (5)$$

with $2 < \alpha < 3$. The detailed analytical and numerical solution of the back-calculation model in 1 given the incubation-period distribution 5 above and the various choices for AIDS growth in 2, 3 and 4 are provided (Comiskey, 1991).

Results

In the above equation, $t = 1$ corresponds to the first year (1982) that AIDS cases were reported in Ireland. The functions 2 to 4, which describe the rate of appearance of new AIDS cases given different growth patterns, were fitted to the AIDS incidence data for 1982 to 1991. The parameter estimates for the three choices of *a(t)* along with r^2, the coefficient of determination, are given in Table 1. The expected annual AIDS incidence these parameters give rise to is given in Table 2.

Solving equation (1) for *h(t)* (Comiskey, 1991) gives rise to the estimates in Table 3 for the incidence of HIV infection in Ireland from 1982 to 1991.

Discussion

The exact solution of the back-projection equation 1 together with the three different growth rates in annual AIDS incidence and the gamma incubation-period distribution 5 provides predictions of 1 351 to 3 054 HIV positive cases in Ireland from 1 January 1982 to 1 January 1990. This is between 1.5 and 3.4 times the number known to the health authorities at that time.

Table 1: Parameter estimates obtained for *a(t)* given three different growth rates in AIDS cases

1.	Quadratic exponential	$a_0 = 0.6065$	$a_1 = 0.5838$	$a_2 = 0.0078$	$r^2 = 93.5\%$
2.	Linear logistic	$b_0 = 0.1514$	$b_1 = 0.0144$	$b_2 = 5.0640$	$r^2 = 94.3\%$
3.	Quadratic	$c_0 = 1.9500$	$c_1 = 2.8140$	$c_2 = 1.0189$	$r^2 = 97.3\%$

Table 2: Expected and observed annual AIDS incidence

Year	Growth 1	Growth 2	Growth 3	Observed
1982	1.08	5.20	0.15	2
1983	1.89	5.40	0.40	1
1984	3.26	5.86	2.68	3
1985	5.53	6.95	6.70	5
1986	9.25	9.41	13.35	6
1987	15.23	14.74	21.75	20
1988	24.69	25.58	32.18	38
1989	39.40	45.07	44.65	51
1990	61.91	70.97	59.15	61
1991	95.77	67.82	75.70	70

Table 3: Estimated incidence of HIV, given different growth rates in AIDS cases

Year	Growth 1	Growth 2	Growth 3	Known
1982	46	26	125	–
1983	71	43	128	–
1984	115	86	138	–
1985	182	181	153	306
1986	287	369	170	229
1987	441	692	190	143
1988	669	1 049	211	116
1989	989	608	236	116
1990	1 470	–4 427	262	110
1991	2 113	–28 544	291	136
Total to				
1 Jan 1990	2 800	3 054	1 351	910
1 Jan 1992	6 383	–	1 904	1 156

Note: – indicates data not available

If the predictions are extended to 1 January 1992, the model predicts between 1 904 and 6 383 HIV positives in Ireland up to that time. This estimate is between 1.6 and 5.5 times the number known at that time to the Department of Health arising from tests carried out at the National Virus Reference Laboratory at University College Dublin.

Looking at the individual predictions more closely, growth rate 1 estimates a total of 2 800 HIV positives up to 1 January 1990. This would appear to be a reasonable prediction given that there were 910 known HIV positives up to that time. However,

after 1990 the model estimates a rapid increase in the numbers testing HIV positive. This is due to the fact that growth rate 1 describes a rapid increase in annual AIDS incidence from that time. This is then reflected in the estimates for HIV incidence.

Growth rate 2 describes increasing incidence in annual AIDS cases up to 1990 followed by decreasing incidence from this time. Estimates of the numbers of HIV positive cases reflect this, with a total of 3 054 positives predicted between 1 January 1982 and 1 January 1990. This model then predicts negative incidence from 1990 and cannot be used to provide estimates for later years.

Growth rate 3 describes increasing annual AIDS incidence up to 1 January 1990 then continues with a steady increase from that time to 1992. Estimates based on this choice of growth rate reflect this increase, with 1 351 cases being estimated for 1982 to 1989 inclusive and 1 904 cases estimated for 1982 to 1991 inclusive.

Overall estimates for the cumulative HIV incidence from 1 January 1982 to 1 January 1992 vary by a factor of 3.35 (1 904 cases to 6 383 cases). This is a considerable improvement on previous estimates. Similar results have arisen in the analysis of the UK data, where estimates varied by a factor of 3 (20 000 to 60 000 cases) (Isham, 1988). In his analysis of these estimates Cox believed the lower end of the range to be more plausible as it corresponded closely to that derived from direct estimation (Cox, 1988).

Bailey discusses the importance of validating results when modelling public-health, and in particular HIV and AIDS, data (Bailey, 1994). One approach he advocates is the cross-checking of findings with independent research results. Some very encouraging results can be obtained if the results discussed above are compared with those arising from transmission models of HIV infections and AIDS in Ireland. These alternative models predict a total of 4 034 infectious people in Ireland in 1991. This number derived from the transmission-model predictions is in the middle of the range predicted by the back-projection method (Comiskey, 1991).

For planning purposes it would seem prudent then to work with an Irish estimate of approximately 2 000 to 4 000 HIV positives in Ireland between 1982 and 1991 inclusive. However, it must be stressed that as yet the exact incubation period is unknown as is how the growth in annual AIDS incidence is changing. All estimates must be viewed with caution until further research into these areas is undertaken. As knowledge of these unknowns increases, the methods discussed in this chapter for estimating the incidence of HIV infection can be applied with increasing confidence.

Model development and applications

To date, the back-calculation model has been developed primarily for estimating the hidden HIV incidence and future short-term predictions of AIDS incidence within a community or within a section of a community as illustrated in the example. As growth in the numbers acquiring disease in one country may follow a different pattern to that of another country, detailed methods developed for one application may need

considerable technical refinement and alteration when applied in a different setting. Research has shown that estimates derived by the back-calculation method are highly sensitive to the choice of AIDS incubation-period distribution used and, as such, much work has been and is still being carried out on improving and refining knowledge of this incubation period.

It is the opinion of the authors that advantage should be taken of the extensive European and global research that has been conducted on the back-calculation method.

These models have not to the authors' knowledge been used in other drug-related areas, in particular to the problem of estimating incidence of drug use within a community. To provide a range of estimates of the prevalence of intravenous drug use in Los Angeles, CA, a combination of methods including multiple capture–recapture, synthetic estimation and back-calculation of AIDS incidence has been used (Hser, 1993). Hser, who is based at the Drug Abuse Research Centre at the University of California, Los Angeles, advocates the employment of complementary methods and data sources.

The back-calculation model has been widely used by the applied statistical and mathematical profession. It has not, however, been widely embraced by the general medical, epidemiological and health professions. This is not due to a lack of data, but more perhaps to the technical nature of the method. Research publications tend to stress developments in the technicalities of the method rather than the relevance of the method to policy-makers. The application of the method to disease prediction was also relatively novel prior to the rapid growth in the number of AIDS cases in the mid-to-late 1980s. There are some interesting exceptions to this: Cox (1988) and Day (1996) both provide predictions of the number of new AIDS cases for the Department of Health in England and Wales based on the back-calculation method. Whether the predictions provided in their reports directly affected and influenced government health policy and resource allocation is a question beyond the scope of this chapter.

Critical assessment of the method for future research

The back-calculation model for estimating the hidden HIV population and AIDS prediction has been applied throughout Europe. The technicalities of the method have been extensively researched and published. It has not to date been used as a means for estimating the hidden drug-using population. Advantage should be taken of developments in the method which uses the known number of AIDS cases (adjusted for reporting delays) to estimate the unknown number of HIV cases. This is accomplished by back calculating from the AIDS cases through the use of the incubation-period distribution to estimate the pool of HIV cases these AIDS cases must have come from.

The same approach may be used to back calculate from the known number of drug users to estimate the larger pool of unknown drug users these known users came from. Consider the case where information is known about opiate users first attending for treatment. (These are analogous to AIDS cases who attend for treatment.) Information is sought from the opiate user on when they first started taking opiate drugs. From this information, the distribution of times from first opiate use to first attending treatment

may be estimated. (This idea is analogous to the distribution of incubation times from first acquiring HIV to developing AIDS.) It is now possible to back calculate from the known numbers of opiate users through the use of the so-called incubation period to estimate the pool of hidden opiate users the known users came from. (This is directly analogous to back calculating from AIDS cases to HIV cases.) The data demands for such research to be conducted are not extensive and in many cases the data exist and are already in place. Many cities have been collecting survey questions on the nature of primary and secondary drug use, age of first use and dates of first treatment for this drug use. The data derived from the answers to all of these epidemiological questions are perhaps under-used at present. Further use of these and other epidemiological studies across Europe would benefit all.

The added value of using the back-calculation method is that within many of the European cities with which the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) has existing links for either capture–recapture or dynamic-modelling studies there already exists the back-calculation technical expertise. In addition, the EMCDDA has access to the relevant drug-treatment data.

Conclusions

In order to apply this method to the problem of estimating drug incidence it is first essential that research into the nature of distribution of times from first misusing drugs to first seeking treatment is carried out. This is analogous to a disease-incubation-period distribution, that is, times from first acquiring infection to first presenting for disease treatment. The data for this study may be drawn from existing local surveys on drug use and first treatment. Very little is known about this so-called incubation-period distribution, yet the back-calculation literature demonstrates that it is precisely the choice of this distribution that will affect future predicted estimates. Depending on the priorities of the research, this study could concentrate on one type or one mode of drug misuse, for example opiate misuse or injecting drug use. One such study is currently being piloted in several European countries under the direction of the EMCDDA (private communication). In addition, a study of the distribution of incubation times or times to first treatment and a subsequent application of the back-calculation method to estimating incidence of opiate use in Dublin, Ireland, has been conducted (Hand, 1999).

Second, a pilot study to assess the feasibility of applying the back-calculation method to estimating the hidden prevalence of drug misuse is being undertaken. This second objective will follow on from and implement the findings on distribution of times to first treatment. To determine the usefulness and reliability of the back-calculation estimates the results should be compared with estimates derived from other methods previously applied. The authors and many other researchers collaborating with the EMCDDA are currently working on prevalence estimates from the multiple capture–recapture method. Given time and economic resource constraints it would seem obvious therefore to compare estimates derived from back calculation with estimates from these soon-to-be-completed studies.

Over a longer time frame, if the results of a back-calculation feasibility study proved worthwhile, the method may be developed further for use with a range of drug-related problems other than HIV and AIDS. If the EMCDDA's back-calculation feasibility study did not produce worthwhile estimates when compared with other methods, then the reports to the EMCDDA will highlight why this was the case, be it deficiencies in the primary data collection or in the variability of the method, and improvements can be recommended. Future surveys and data collection can then take note of these recommendations.

With regard to the resource needs of such a feasibility study, many of the data structures are already in place. The technicalities of the method are also well documented. There would be requirements for access to the necessary epidemiological data and their appropriate collation. Finally, for such a study to be of benefit to the wider community, it is essential that the implications of the research and estimates derived be placed within the context of relevance and implication for health professionals, policy-makers and those with the responsibility of scarce and valuable resource allocation.

In conclusion, to quote the pioneering biomathematician N. T. J. Bailey (1975):

> *In the face of misery and suffering on a monumental scale, epidemic theory is a luxury mankind can ill afford. The world must not only be interpreted: it must be changed.*

References

Anderson, R. M. (1989). 'Mathematical and statistical studies of HIV'. *AIDS* 3: 333–46.

Anderson, R. M., Medley, G. (1988). 'Epidemiology, HIV infection and AIDS: the incubation and infection periods, survival and vertical transmission'. *AIDS* 2 (suppl. 1): S57–S64.

Bailey, N. T. J. (1994). 'Estimating HIV incidence and AIDS projections. Prediction and validation in the public health modelling of HIV/AIDS'. *Statistics in Medicine* 13: 1933–43.

Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications*. London: Griffin and Co. Ltd. Second edition (first edition 1957).

Becker, N. G., Chao, X. (1994). 'Dependent HIV incidences in back projection of AIDS incidence data'. *Statistics in Medicine* 13: 1945–58.

Brookmeyer, R., Damiano, A. (1989). 'Statistical methods for short-term projections of AIDS incidence'. *Statistics in Medicine* 8: 23–34.

Brookmeyer, R., Gail, M. (1986). 'Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States'. *Lancet* 2: 1320–22.

Comiskey, C. M. (1996). 'Integral equations and estimating the incidence of HIV infection'. *Zeitschrift Fur Angewandte Mathematik Und Mechanik* 76 (S2): 505–6.

Comiskey, C. M. (1992). 'Improvements in integral equation models for estimates of the level of HIV infection in Ireland'. *The Statistician* 41: 329.

Comiskey, C. M. (1991). 'Modelling the transmission dynamics of HIV and its progression to AIDS in Ireland'. PhD thesis, Dublin City University, Ireland.

Comiskey, C. M., Ruskin, H. (1992). 'AIDS in Ireland. The reporting delay distribution and the implementation of integral equation models'. *Computer Application in the Biosciences* 8 (6): 579–81.

Cooley, P. C., Myers, L. E., Hamill, D. N. (1996). 'A meta analysis of estimates of the AIDS incubation distribution'. *European Journal of Epidemiology* 12: 29–35.

Cox, D. (1988) *Short-term prediction of HIV infection and AIDS in England and Wales*. London: Her Majesty's Stationery Office.

Day, N. E. (1996). 'The incidence and prevalence of AIDS and prevalence of other severe HIV disease in England and Wales for 1995 to 1999: projections using data to the end of 1994'. *Communicable Disease Report Review* 6 (1): 1–8.

Day, N. E., Gore, S. M. (1988). 'Predictions of the number of new AIDS cases and the number of new persons infected with HIV up to 1992: the results of "back-projection" methods', in Cox, D. *Short-term prediction of HIV infection and AIDS in England and Wales*, 74–8.

Day, N. E., Gore, S. M., McGee, M. A., South, M. (1989). 'Predictions of the AIDS epidemic in the UK: the use of the back-projection method'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 325: 123–34.

Dietz, K., Seydel, J., Schwartländer, B. (1994). 'Back-projection of German AIDS data using information on dates of tests'. *Statistics in Medicine* 13: 1991–2008.

Downs, A. M., Ancelle, R. A., Jager, H. J. C., Brunet, J. B. (1987). 'AIDS in Europe: current trends and short-term predictions estimated from surveillance data'. *AIDS* 1: 53–7.

Downs, A. M., Heisterkamp, S. H., Brunet, J. B., Hamers, F. F. (1997). '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* 11: 649–62.

Hand, E. (1999) 'Estimating the prevalence and the latency period of opiate use in Dublin'. Thesis (MSc), National University of Ireland, Maynooth.

Hendricks, J. C. M., Medley, G. F., Heisterkamp, S. H., Vangriensven, G. J. P., Bindels, P. J. E., Coutino, R. A., Vandruten, J. A. M. (1992). 'Short-term predictions of HIV prevalence and AIDS incidence'. *Epidemiology and Infection* 109: 149–60.

Hser, Y. I. (1993). 'Population estimates of intravenous drug users and HIV infection in Los Angeles County'. *International Journal of the Addictions* 28: 695–709.

Isham, V. (1988). 'Estimation of the incidence of HIV infection – the back-projection method', in Cox, D. *Short-term prediction of HIV infection and AIDS in England and Wales*, 53–55.

Mariotti, S., Cascioli, R. (1996). 'Sources of uncertainty in estimating HIV infection rates by back calculation. An application to Italian data'. *Statistics in Medicine* 15: 2667–87.

Marschner, I. C. (1994). 'Using time of first positive HIV test and other auxiliary data in back projection of AIDS incidence'. *Statistics in Medicine* 13: 1959–74.

Rosenburg, P. S. (1994). 'Back-calculation models of age-specific HIV incidence rates'. *Statistics in Medicine* 13: 1975–90.

Wilson, S. R., Fazekas de St. Groth, C., Solomon, P. J. (1992). 'Sensitivity analyses for the back-calculation method of AIDS projections'. *Journal of Acquired Immune Deficiency Syndrome* 5: 523–7.

CHAPTER 7

COMPARTMENTAL MODELLING AND STOCHASTIC DYNAMIC SYSTEMS

Gordon Hay and Catherine Comiskey

Many quantities are of interest to drug-misuse researchers, such as the number of people that inject heroin or the number of children that begin using drugs at a particular age. These two quantifiable 'groups' are examples of compartments. The number of people in each of these groups, or compartments, may be dependent on the number in others, for example the number of people that inject heroin may be related to the number that smoke the drug, whereas the number of children who begin using drugs may be related to the number that smoke or drink alcohol.

Compartmental models provide a framework in which numbers of people in different compartments, and the relationships between different compartments, can be described analytically. Two different types of compartmental models are described in this chapter: deterministic models in which the numbers in each category are completely determined by the model; and stochastic models in which random effects can result in different outcomes.

Compartmental models are well established in modelling the spread of diseases, including HIV/AIDS. Recent models which examine the spread of HIV in drug-using populations split those populations into different compartments. Although these compartments are created to make disease-spread models more realistic, for example by assigning drug injectors into groups that share needles daily, weekly or monthly, drug-using populations can be described in other ways, by examining drug-using careers or levels of drug use. Drug-related consequences can be added into compartmental models, such as social problems or health-care-related costs.

Some other examples of compartmental models, in particular some of the few applications to drug-related problems, are considered in this chapter. The use of these modelling techniques is then assessed and recommendations are made on the feasibility of the future use of these models by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA).

Figure 1: Deterministic and stochastic models

Two main types of mathematical models have commonly been used to describe the spread of diseases.

Deterministic models are easy to formulate and easy to solve, however such models do not include random fluctuations and therefore may not be suitable for modelling some epidemiological processes, particularly when the numbers of people infected are small.

Stochastic models include random fluctuations in the numbers of people infected, although these models are often harder to work with. Computer programs can, however, assist in working with stochastic models.

Both types of model assume that the population can be split into compartments, and the result from the models is often the number of people infected over time.

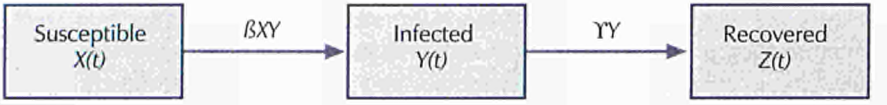
Model description

This section addresses how mathematical models can be developed to describe the spread of a disease such as HIV/AIDS. Although there are two schools of thought in epidemiological mathematical modelling (see Figure 1), both assume that the population in which the disease spreads can be grouped into compartments, such as those that are infected and those that are not. One type of mathematical model, the deterministic model, assumes that the numbers infected increases or decreases at a specific rate. This rate is usually related to the numbers in different compartments and the rate can change over time. In a stochastic model, the rates at which the number infected changes is replaced by probabilities, thus the random nature of disease spread can be included. Once the population has been split into relevant compartments, it is a comparatively easy exercise to describe mathematically how the size of these compartments will change over time. For example, in the deterministic model this is done by formulating a set of differential equations. These equations can easily be solved to give the number of people infected over time. The stochastic equivalents are harder to solve, and computational methods are often needed to provide results.

As a necessary precursor to using either modelling technique, a way of describing the transmission dynamics of the particular disease using numbers has to be created. This is initially done in the context of a deterministic model which is often easier to work with. Compartments can be constructed, and when describing the spread of a disease such as measles three compartments can be constructed to describe the different stages in the disease process: susceptible; infected; and recovered (see Figure 2).

The terms 'susceptible' and 'infected' are self-explanatory; however, it should be noted that as infection with measles results in immunity, the people in that group cannot be reinfected. This basic description is often called the SIR model, and each individual in a population will be in one of these three compartments. At each point in time, there will be a fixed number of people in each of these compartments, and using mathematical notation, the numbers of people in the three compartments can be called $X(t)$, $Y(t)$ and

Figure 2: The SIR model



$Z(t)$. The presence of t in the notation is a reminder that the number of people in each compartment changes over time.

This change can be described in a system of differential equations. In this system, the rate of change of each variable, denoted by $dX(t)/dt$, $dY(t)/dt$ and $dZ(t)/dt$, depends on the values $X(t)$, $Y(t)$ and $Z(t)$, as shown in equation 1.

$$\frac{dX}{dt} = -\beta XY \tag{1}$$

$$\frac{dY}{dt} = \beta XY - \gamma Y$$

$$\frac{dZ}{dt} = \gamma Y$$

Here, by convention, $X(t)$ in the differential equation is replaced by X for convenience, although the value is still dependent on the time t. There are several things of note in this system of differential equations, which is a deterministic model for the spread of measles. First there are two parameters in this model, β and γ. The first is the infection rate and the second the recovery rate. These are the only parameters in the model and as such the model is very simple as it does not include other processes such as birth of susceptibles or removal from any compartment for reasons such as death. If total population size is denoted as $N(t)$, then $N(t) = X(t) + Y(t) + Z(t)$, or more interestingly

$$\frac{dN}{dt} = \frac{dX}{dt} + \frac{dY}{dt} + \frac{dX}{dt} \tag{2}$$

$$= -\beta XY + \beta XY - \gamma Y + \gamma Y$$

$$= 0$$

which shows that the population size is fixed. What should also be noted from equation 1 is that the rate at which the number of susceptibles decreases is proportional to the product of X and Y. In other words, the number of new infections depends on the contact rate between susceptibles and infectives and this is known as the mass-action law.

As previously stated, this is a very basic model and there have been refinements to it in the context of measles modelling, and also in modelling the spread of other diseases. One further example is a model that examines the spread of a sexually transmitted disease. Here the population of males and females must be described separately. If the number of susceptible, infected and recovered males is given as $X(t)$, $Y(t)$ and $Z(t)$, and the number of susceptible, infected and recovered females as $x(t)$, $y(t)$ and $z(t)$, then a basic model, which is just an extension of the previous model to two interacting populations, will be

$$\frac{dX}{dt} = -\beta Xy \tag{3}$$

$$\frac{dY}{dt} = \beta XY - \gamma Y$$

$$\frac{dZ}{dt} = \gamma Y$$

$$\frac{dx}{dt} = -\beta' xY$$

$$\frac{dy}{dt} = \beta' xY - \gamma' y$$

$$\frac{dz}{dt} = \gamma' y$$

Here there will be two infection rates, β for males becoming infected by infectious females and β' for females becoming infected by males. Similarly there will be two recovery rates γ and γ' for males and females respectively. Again the mass-action law is relevant as the number of new infections in males is proportional to the product of X and y and in females is proportional to the product of x and Y.

In the context of deterministic modelling, these basic models can be simply extended by including either the number of interacting populations or the number of compartments that describe a single population. A good example of this type of modelling is that which describes the heterosexual spread of HIV. While there are still two interacting populations, males and females, the number of compartments which describe the spread of the disease can be increased to include an AIDS compartment, as well as a compartment which contains those who are infected but not yet infectious. Another way of extending these simple deterministic models would be to include more interacting populations. For example, the homosexual spread and the needle-sharing spread of HIV can all be modelled simultaneously by constructing compartments for each population and modelling the rate at which the numbers in these compartments change between populations. A good example of this would be female drug injectors who are also prostitutes and thus come into contact with male heterosexual clients.

These deterministic models are quite simple to construct and computational methods have been developed to solve the differential equations to show how the number of infected people changes over time. However, as the name suggests, the spread of the disease is determined at the start of the epidemic in that there is no randomness in the disease process. Thus deterministic models can be seen as an approximation to more realistic stochastic models in which randomness is included. Unless the populations that are being examined are small, or the numbers in some of the compartments as described above become small, then deterministic models are often adequate.

To include randomness, stochastic models examine the probability that a person changes from one compartment to another, instead of assuming that the numbers change at a determined rate. Equation 1 can be converted into a corresponding stochastic model as in equation 3 where, only looking at the numbers of susceptibles and infectives, in a small time period Δt

$$\Pr\{(X,Y)\rightarrow(X-1,Y+1)\}=\beta XY\Delta t \qquad (4)$$

$$\Pr\{(X,Y)\rightarrow(X,Y-1)\}=\gamma Y\Delta t$$

In other words, the probability that one susceptible becomes infected within the small time period Δt is $\beta XY\Delta t$ and, if this happens, the number of susceptibles decreases from X to $X-1$ and correspondingly the number of infectives increases from Y to $Y+1$. In constructing these probabilities it is assumed that only one event can happen in the small time interval Δt. As in the deterministic model, the probability is proportional to the product of the number of susceptibles and infectives.

Stochastic models can also be extended by including more compartments and interacting populations, but the theory behind stochastic models is far more complicated than that behind deterministic models and thus stochastic models in which solutions to these types of equations are examined are rare.

A more common form of stochastic modelling which employs compartments is the use of Monte Carlo simulation models. In these the probability of an event, such as a susceptible becoming infected, is evaluated within a computer program and a random number generator is used to decide whether or not this event should be simulated. Thus approximations to stochastic models can be readily evaluated and stochastic simulation languages are becoming more and more popular to simulate processes within science and engineering.

While the examples above, like most compartmental models, describe a time-dependent process, this does not always have to be the case. In some diseases, the age of a person is important, therefore the number of susceptibles, infectives and recovereds may be better described as $X(a,t)$, $Y(a,t)$ and $Z(a,t)$ where age is denoted by a. The numbers in these compartments can vary over time and over age groups and differential equations can be constructed to describe the process. The parameters in these models are often age-dependent. These differential equations are more complex than equations 1 and 2 as they involve the rate of change over time and over age and are more correctly known as partial differential equations. They can be solved easily using numerical

methods and the numbers of people infected at different times and at different ages can be calculated.

Just as the models which only examine the rate of change over time can be extended to model the rate of change between ages, deterministic models can also be extended to include spatial effects. As previously described, there can be two interacting populations such as males and females in which compartments can be created to describe the sexual spread of a disease. This is similar to a situation in which there are two or more towns with the disease spreading between them. Here, a compartmental structure with differential equations can be created including both disease spread within towns and between towns.

Where have the models been applied?

Compartmental models, and the deterministic or stochastic methods used in conjunction with them, have many uses, in particular biological population dynamics and epidemiology. The use of mathematics in describing social processes such as the spread of disease has been well established since the start of the twentieth century when Hamer (1906) examined the spread of measles and Ross (1911) examined the spread of malaria. Although the fields of epidemiology and mathematical modelling have advanced rapidly since then, some assumptions and principles are basic to subsequent research. Advances in modelling sexually transmitted diseases have been made, for example Hethcote and Yorke (1984). Anderson and May (1991) provide an excellent review of mathematical modelling of infectious diseases. The growth of animal populations and the interaction between different populations have been studied extensively, including predator–prey models in which the population dynamics of one animal depend on the numbers of another animal on which they prey. Renshaw (1991) provides a review of these fields, in which he also examines spatial population dynamics.

Kaplan (1989) used a deterministic model to describe the spread of HIV among drug injectors. In this basic model he assumed a constant population size and that the disease spread through drug injectors sharing needles in shooting galleries. As such, this model was not very realistic in predicting the future course of the disease; however, the use of control strategies, such as needle exchanges, was evaluated. Indeed, it was a condition of the legalisation of needle exchanges in the US state of Connecticut that some kind of evaluation of a needle exchange was undertaken. Caulkins and Kaplan (1991) extended this model to allow for a varying population size of drug injectors. In this model they showed the population dynamics of the intravenous (IV) drug-using population as

$$\frac{DN(t)}{dt} = cN(t)^{\nu} - \mu N(t) \tag{5}$$

Here they assume that the rate at which people join the population of IV drug users is proportional to the number of current users raised to some power ν, where ν is non-negative and strictly less than one. They present two arguments about the potential value of ν. If the new users begin injecting because of other people injecting then the

number of new injectors increases with the number of current injectors and hence ν would be close to one. If, however, injectable drugs are freely available such that any person who is predisposed to inject drugs will and the recruitment rate is determined by social factors then ν could be quite small. At the extreme, if ν is zero then people would begin injecting drugs at a constant rate c. Caulkins and Kaplan also assume that current injectors cease injecting at a constant rate μ. They go on to estimate some of the parameters used in the model, both before and after the introduction of AIDS, and show that due to AIDS, the number of drug injectors will decrease.

This population-size model is combined with a disease-spread model in which a population of drug injectors shares possibly infected needles. Thus three compartments are modelled, $\beta(t)$, the fraction of needles that are infected, $I(t)$ the number of infected injectors and $U(t)$ the number of uninfected injectors at time t. Differential equations which describe the transmission dynamics are constructed and employ the parameters related to population size and to the infection process.

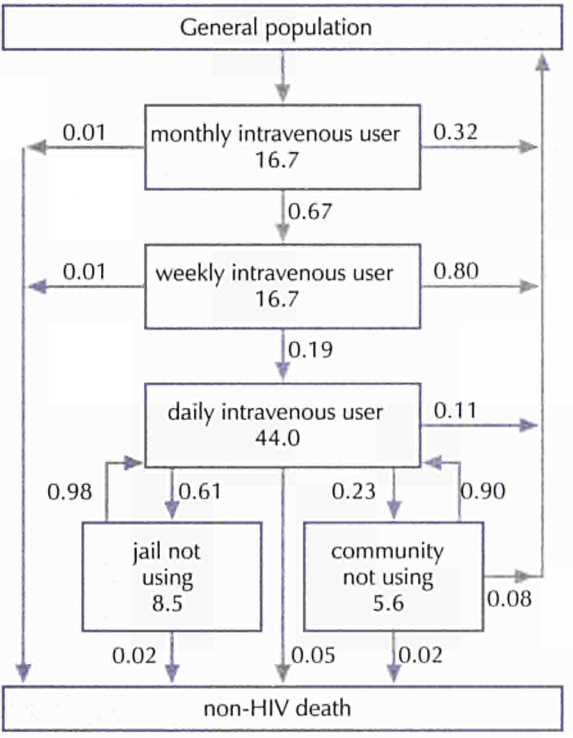
Billard and Dayananda (1993) look closer at the rate at which people begin using drugs in their eloquently titled paper 'Drug addiction – pushers generated from the addicts'. They combine a population-size deterministic model with a previous model of Hoppensteadt and Murray (1981) who considered the concentration of a drug in the blood of drug users. Billard and Dayananda use the mass-action law in that the rate of change of the number of addicts is proportional to the product of the number of addicts and non-addicts. They also assume that pushers are normally hard-core drug users such that the number of pushers will be proportional to the number of addicts. This highly mathematical paper is not at all realistic in the context of predicting how many people will begin using drugs, but demonstrates some of the theory behind deterministic and stochastic modelling.

Some models that examine the spread of HIV among drug injectors provide basic attempts at describing the population dynamics of drug use. Peterson et al. (1990) use Monte Carlo simulation methods to model the spread of HIV among drug users. They use six compartments to model the progression of HIV, ranging from susceptible to death from AIDS. They also create a parallel system which describes the career progression of drug injectors. Figure 3, which is adapted from Peterson's paper, presents the compartments that are used to describe the career progression of drug users.

Several things are worth noting from this diagram. Although Peterson goes on to model this scenario with stochastic methods, his compartmental framework can quite easily be modelled using deterministic methods. In the figure, he and his colleagues chose to present values for the probabilities instead of working with parameters such as β or γ.

In the 'Daily IV user' compartment, the mean duration in months assumed for this compartment is 44.0. In other words, once a person begins injecting daily, they will, on average, remain injecting daily for 44 months. Certain events can happen to a daily drug injector: they can end up in jail; end up in a community treatment programme; cease injecting completely and therefore go back to being a member of the general population; or die from something unrelated to HIV. It should be remembered that a parallel compartmental modelling system describes the disease progression and this

Figure 3: Compartments used to model the career progression of drug users



Source: adapted from Peterson et al. (1990)

system would include death from AIDS. Each event has a probability attached; for example the probability that a daily drug injector will go to jail is 0.61.

As can be seen from Figure 3, this compartmental system relies on probabilities that can perhaps be estimated from sociological research. For example, studies can look at the probability of drug injectors ceasing injecting and can also look at the non-HIV-related mortality of drug injectors. In other compartmental models, some of the probabilities are harder to estimate. For example in disease models very little is known about the probability of becoming infected with HIV after using a needle which is contaminated with the virus.

Other compartmental models have been used to model the spread of HIV among drug injectors, either in isolation (Greenhalgh and Hay, 1997) or in conjunction with sexual-spread models (Blower et al., 1991; Arcá et al., 1992). However, these models do not specifically look at changes in the size of the drug-injecting population.

Another type of stochastic model which has been used to model the spread of disease is the chain-binomial model. In this type of model, the number of new infections caused by an infected person is assumed to follow a binomial distribution. For example, an infective will infect one or more people with different probability and these people will then go on to form the next link in the chain by infecting other susceptibles. Mackintosh and Stewart (1979) use this technique to model a heroin epidemic, as did de Alarcón (1969), who demonstrated this technique using data collected from a community. This type of modelling assumes that heroin use follows contact with other users and can model the scenario in which drug use is introduced fairly rapidly into a community.

Model development

A good example of model development is that of the models to describe the homosexual spread of HIV. The initial models did not differ much from those which described the spread of diseases such as measles except that there was not a 'recovered' class in HIV models. However, it soon became clear that the transmission dynamics of this disease are more complex, particularly because a new infection results from a sexual act, and this could not be modelled in the same way as a disease like measles which is airborne. The probability of becoming infected would depend on how sexually active an individual was and thus the mass-action law may not be strictly relevant. The selection of sexual partners can also be included in these models as the probability of becoming infected would be greater for someone who has intercourse with many people than a person who is in a monogamous relationship. The use of condoms will also affect the spread of the disease, as would biological parameters such as the infectivity of the virus at different stages of the disease progression. Thus what originated as a very simple model quickly became a complex model which realistically described the transmission dynamics of a particular disease.

It may therefore be possible to develop the basic SIR model to make it relevant to drug-related areas. In order to model non-drug users as a susceptible compartment and drug users as an infective compartment, the process that makes someone begin using drugs must be examined. A rate which describes people starting to use drugs can be assigned, and this rate may depend on the number of others in the population using drugs as in the mass-action law, or it may depend on other factors. As in AIDS models, drug use can also be split into more than one compartment, either by severity of drug use or by type of drug use.

It may be possible to develop models to describe the start of drug use by young people. Research by Barnard et al. (1996) shows increasing numbers of children smoking, drinking and abusing solvents as they progress through school. What is of interest from that research is that the number of children smoking at age 12 is similar to those drinking at 13, those using solvents at 14 and those using drugs at 15. Whether or not these activities are related is an exercise in sociological research; however, these different events and the processes linking them would be ideal to describe as a compartmental model, especially as research involving young people often examines the age at which children begin using licit or illicit substances.

Critical assessment of model types for future use

Various types of mathematical models have been presented above which attempt to describe a sociological or medical process, such as the transmission dynamics of HIV in a population of drug injectors, by creating compartments. Some processes are more amenable to being described by compartments, such as the 'susceptible', 'infected' and 'removed' compartments which describe the spread of measles. Other processes, such as the more sociological process of a drug-using career, may be less suited to these techniques.

For example, does the probability that someone begins using drugs depend on the number of others within the population who are using drugs, or is drug use a consequence of growing up such that a certain proportion of young people will begin using drugs regardless of others? Is drug use such a rare occurrence that it can only begin after contact with another drug user, thus making a chain-binomial model more relevant?

There could be other influences on the number of people using drugs, such as the relative success or failure of a drug-prevention campaign, or the availability of drugs. The latter could be modelled as a separate compartmental system and this would raise questions relating to what factors affect availability, such as supply and demand or law-enforcement activities. This type of modelling would perhaps then be more akin to economic models. Even if it was concluded that the number of people using drugs, or the number of people who begin using drugs, depends on availability, what would be the parameters in a compartmental system that would describe them and how would these values be measured?

The perennial question of definitions may need to be raised in creating compartments to describe, for example, a drug-using career. It would be difficult to model the progression from user to problem user to addict without a clear consensus about these definitions. In assigning an 'addict' compartment it is easy to model some transitions out of this compartment such as death, but what would be the process undertaken by a person to move from 'addict' back to 'user'? An issue related to definitions would be the different drugs used. In some areas and at different times the use of stimulants may be seen as a serious problem and modelling the use of these drugs may be of interest, whereas in other areas or at other times the use of these drugs may be seen as a stage in a process that results in opiate addiction.

Although compartmental models may be better thought of as tools to examine 'what if' scenarios, there is still a need to employ realistic data where available. Blower and Medley (1992) describe the application of mathematical modelling techniques, including compartmental models, to examining the spread of HIV among drug injectors. Instead of examining the advances in the methodology, they discuss the more relevant issue of the availability of data that would be of use in creating models and the usefulness of this data. They detail problems with existing data, for example questionnaires that code parameters of interest such as the number of people who have shared injecting equipment within the last six months into categories such as 0–5, 5–10. Such questions would not be sensitive to small behavioural changes, and as the responses to such

variables are usually very skewed, it would be of more use to have single numbers as categories at the lower end of the scale.

Unlike mathematical models of infectious-disease spread which employ biological parameters, the sociological parameters that would be needed to model drug-related problems may be more transient. For example, the length of time that someone remains infectious with a certain disease can be documented and this infectious period, or even an age-specific infectious period, would be common throughout Europe. This would be in contrast to, for example the reasons for a person stopping using drugs in Athens or Rome which may be very different to those for someone in Helsinki or Stockholm.

Recommendations

Although existing research which does not specifically focus on disease spread is limited, compartmental models have an exciting potential for future development. Two main areas of future research are proposed. The first is one in which the 'drug-using career' is quantified and assessed, for example by looking at existing data sources which report the time between a drug user beginning to use drugs and presenting at services. The other main area of research would be to examine children's drug use, based on studies that show that young people may begin using drugs after they begin smoking or drinking. More specifically, the following is recommended.

Examining the feasibility of using epidemic models to model drug use as a disease

While it is relatively easy to create compartments with which to describe different stages, the more pertinent questions relate to the processes that link the compartments, such as the reason for someone who already uses drugs to begin injecting, and the estimation of such parameters. It may be possible to answer these questions by undertaking research into these processes; however, another approach would be to review the available data, such as those collected by the Pompidou Group of the Council of Europe, and seek to adapt them for inclusion in a mathematical model. Such a review may identify problems with the existing data, or with data that are not routinely collected but may be readily available. It may be possible to glean data from the existing literature; for example, models which describe the spread of HIV through shared injecting equipment require an estimate for the probability that injecting with an infected needle will result in a new infection. This probability is often estimated from studies of needle-stick injuries; however, Kaplan (1992) demonstrated an interesting use of his mathematical models in which he estimated this parameter.

Within the context of other drug-related problems, research may be able to provide an insight into parameters such as the average time people use drugs before contacting services. Such information would have a dual use in developing back-calculation methods (see Chapter 6) as well as in compartmental models. A review of the literature may also seek to answer some of the questions presented above, for example the progression from one type of drug use to another and the increasing severity of drug use. Where

available, it would be preferable to use comparable data from throughout Europe to ascertain which processes are common to different areas in the same way that biological processes would be, and which processes vary over space and time.

Feasibility of modelling the processes leading up to young people using drugs

The academic literature often suggests a link between young people's licit substance use, such as smoking or drinking, and their drug use such that those who begin drinking or smoking at a young age are more likely to take drugs at a later age. Thus there are at least four compartments with which the underlying processes could be modelled: drinking; smoking; solvent abuse; and drug misuse. The use of drugs other than cannabis may be assigned another compartment, as can a parallel system such as the availability of drugs or alcohol.

As in the models that describe the spread of diseases, the propensity to begin using drugs may be related to the number of the children's peers that use drugs; however, sociological research is yet to confirm that this is a relevant factor. It may then be preferable to model the uptake of drug use as something that happens to a proportion of young people, unrelated to the number of other young people that are using drugs. The previous two scenarios can both be modelled using the formula proposed by Caulkins and Kaplan and, while it may not seem sensible to attempt to estimate the parameters employed in this formula by commissioning sociological research, it may be prudent to adapt the existing knowledge about the factors relating to someone's inititation into drug use to indicate the likely values of these parameters.

Both these recommendations seek to explore sociological processes that have already been identified and examined, and this is reflected in the academic literature, particularly in relation to longitudinal studies. Pertinent questions can therefore be raised about both the applicability and the 'added value' of these methods. To answer these questions requires a definition of epidemiological mathematical modelling, such as that offered by Blower and Medley (1992). Compartmental models, regardless of which theory is used to analyse them, cannot at this time give precise answers to some of the more pressing questions in the field of drug misuse. What they can do is to increase the understanding of the processes and, perhaps more interestingly, illustrate the interactions between processes which may previously have been examined separately. Commenting on some of the longitudinal studies which have examined the social processes described above, namely the factors which determine a young person's uptake of drugs and the switch from smoking to injecting, can demonstrate this point.

In both these situations, examining people over a period of time can, for example, offer great insight into the reasons why that individual has changed their route of administration from smoking or 'chasing the dragon' to injecting. A good example of this research would be the drug transitions study at the UK National Addiction Centre (Strang et al., 1997). As part of this study, the reasons why individuals switch from smoking to injecting was examined and valuable information gathered on an individual's reasons for behavioural change. Similarly, longitudinal studies have proposed reasons why young

people begin using drugs, particularly in the United States (see Newcomb and Bentler, 1986). While these and other studies explore the reasons why individuals make choices related to their drug or substance use, their generality at population level has yet to be evaluated. Within a stochastic compartmental model the variability of a population can be assigned by attaching a distributional form to a parameter or process, although this is still in an attempt to improve the generalisation at population level rather than implicitly to describe each individual. Deterministic models seek to describe these processes by a parameter such as a rate. Moreover, compartmental models seek to simplify the individual's processes, perhaps even to over-simplify these processes. This, however, does not contradict the definitions of Blower and Medley. By describing these processes either by a rate, as in the deterministic models, or a distribution, as in the case of stochastic models, the advantage of compartmental models is their ability to understand the interaction between the different ongoing processes. Thus compartmental models use basic mathematical modelling techniques to describe, or simplify, the individual processes relating to drug misuse. However, the benefits of such simplifications are to enable the different processes to be joined together within a logical framework of compartments, to give an insight into the complete picture. This complete picture may then be the result of a 'thought experiment' as described by Blower and Medley (1992).

Conclusions

To conclude, compartmental models use mathematics to describe not only individual processes, but also how these processes join together or interact to present a more complete picture. The step from examining individual processes to examining the complete picture necessitates some simplification, and this is reflected in the difficulties in using compartmental models to provide precise predictions. Compartmental models can, however, offer great insights into the interaction between these processes, and indeed into situations which are often described as one process, such as the reasons why a young person would begin using drugs, but which may be better attributable to several underlying processes.

Epidemiological drug researchers are aware that the science is necessarily vague. It may never be able to answer precisely questions such as how many people use drugs or how long a person will smoke heroin before moving on to injecting. This does not detract from researchers' interests in these questions; however, it should be recognised that compartmental models may be the best way to explore some of the underlying processes. Of course, compartmental models may not answer all questions about drug misuse as there can never be a single methodology which answers all questions. The uses of this methodology in the situations described above and also in the study of other drug-related problems, should, however, be explored thoroughly.

References

Anderson, R. M., May, R. M. (1991). *Infectious diseases of humans*. Oxford: Oxford University Press.

Arcá, M., Perucci, C. A., Spadea, T. (1992). 'The epidemic dynamics of HIV-1 in Italy: modelling the interaction between intravenous drug users and heterosexual population'. *Statistics in Medicine* 11: 1657–84.

Barnard, M. A., Forsyth, A. J. M., McKeganey, N. (1996). 'Levels of drug use among a sample of Scottish school children'. *Drugs: Education, Prevention and Policy* 3: 81–9.

Billard, L., Dayananda, P. W. A. (1993). 'Drug addiction – pushers generated from the addicts'. *Biometrical Journal* 35: 227–44.

Blower, S., Medley, G. (1992). 'Epidemiology, HIV and drugs: mathematical models and data'. *British Journal of Addiction* 87: 371–9.

Blower, S. M., Hartel, D., Dowlatabadi, H., May, R. M., Anderson, R. M. (1991). 'Drugs, sex and HIV: a mathematical model for New York City'. *Philosophical Transactions of the Royal Society Series B: Biological Sciences* 321: 171–87.

Caulkins, J. P., Kaplan, E. H. (1991). 'AIDS impact on the number of intravenous drug users'. *Interfaces* 21: 50–63.

de Alarcón, R. (1969). 'The spread of heroin abuse in a community'. *Bulletin of Narcotics* 21: 17–21.

Greenhalgh, D., Hay, G. (1997). 'Mathematical modelling of the spread of HIV/AIDS amongst injecting drug users'. *IMA Journal of Mathematics Applied in Medicine and Biology* 14: 11–38.

Hamer, W. H. (1906). 'Epidemic disease in England'. *Lancet* i: 733–9.

Hethcote, H. W., Yorke, J. A. (1984). 'Gonorrhea: transmission dynamics and control'. *Lecture Notes in Biomathematics* 56: 1–105.

Hoppensteadt, F. C., Murray, J. D. (1981). 'Threshold analysis of a drug use epidemic model'. *Mathematical Biosciences* 53: 79–87.

Kaplan, E. H. (1992). 'A model-based estimate of HIV infectivity via needle sharing'. *Journal of Acquired Immune Deficiency Syndromes* 5: 1116–18.

Kaplan, E. H. (1989). 'Needles that kill: modelling human immunodeficiency virus transmission via shared drug injection equipment in shooting galleries'. *Reviews of Infectious Diseases* 11: 289–98.

Mackintosh, D. R., Stewart, G. T. (1979). 'A mathematical model of a heroin epidemic: implications for control policies'. *Journal of Epidemiology and Community Health* 33: 299–304.

Newcomb, M., Bentler, P. (1986). 'Frequency and sequence of drug use. A longitudinal study from early adolescence to young adulthood'. *Journal of Drug Education* 2: 101–20.

Peterson, D., Willard, K., Altmann, M., Gatewood, L., Davidson, G. (1990). 'Monte-Carlo simulation of HIV infection in an intravenous drug user community'. *Journal of Acquired Immune Deficiency Syndromes* 3: 1086–95.

Renshaw, E. (1991). *Modelling biological populations in space and time*. Cambridge: Cambridge University Press.

Ross, R. (1911). *The prevention of malaria*. London: Murray.

Strang, J., Griffiths, P., Powis, B., Abbey J., Gossop, M. (1997). 'How constant is an individual's route of heroin administration? Data from treatment and non-treatment samples'. *Drug and Alcohol Dependence* 46: 115–18.

CHAPTER 8

DYNAMIC MODELS IN DRUG EPIDEMIOLOGY

Colin Taylor and Philip Young

In this chapter the term 'dynamic model' is used to cover methods from system analysis and time-series modelling which are applied to drug-prevalence estimation, but which do not simply attempt to chart or measure prevalence, but rather attempt to model the underlying processes. A useful distinction among such methods is to categorise them as either dynamic or static models. Static methods evaluate a population at a single point in time, while dynamic methods are appropriate when a population is traced over time.

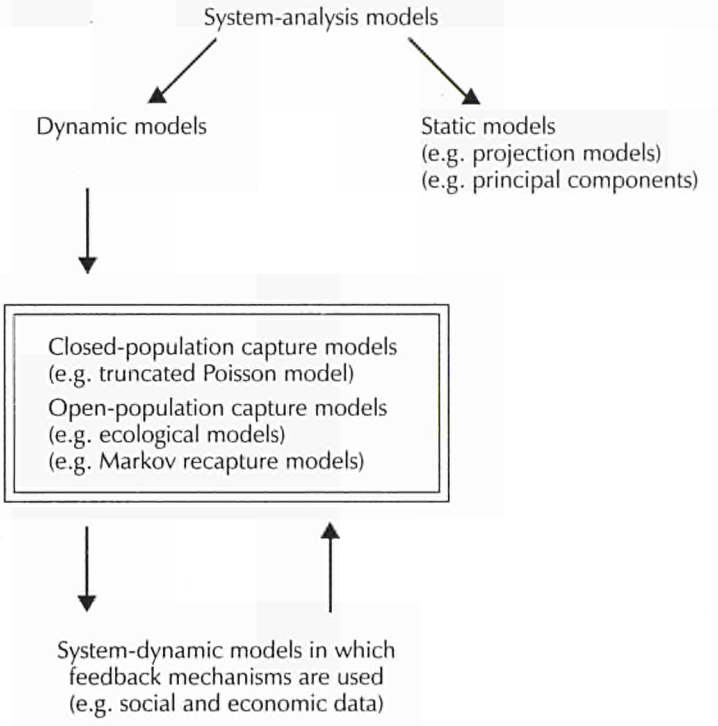
Because of their simplicity, static models have found favour in prevalence estimation. The simplest static-estimation techniques are the synthetic-estimation techniques. These methods develop prevalence estimates for new populations using readily available data sources, or indicators, from known populations by matching various predictor variables, usually demographic characteristics (e.g. ethnicity, gender, age and regional location) and determining appropriate weighting schemes. The appropriate weights are typically determined by either projection models or principal-component analysis.

Dynamic models differ from the static models in that they describe processes over time. The process examined may be the states of individual drug users via multiple-capture models. Multiple-capture models are divided into two classes: open-population; and closed-population models. In a closed-population model it is assumed that the population remains the same throughout the time period, while in an open model adjustments are made for the fact that members of the original population may leave and new members may enter. Techniques for closed-population methods are much simpler, such as the truncated Poisson model, and other than the closed-population assumption make fewer additional assumptions. However, if the time period is sufficient, then closed-population methods will be biased and the more complex open-population models may provide a better estimate of prevalence.

If only aggregate data are available, then a system-dynamic model may be used. Although such models only use aggregate data, they can provide fairly sensitive estimates of prevalence because they typically employ additional economic and social data in a feedback mechanism to adjust the crude aggregate data.

The choice of model selected is largely dictated by the data that are available, and their interrelationships are summarised below in Figure 1.

Figure 1: The inter-relationships of the various types of system-analysis model



Dynamic models

The term 'dynamic model' appears to have no particular specific usage in the drug-research field. Instead it would appear to be open very much to interpretation, carrying with it allusions and references from other disciplines where the usage is more specific. Broadly speaking, within the domain of general drug-prevalence research, it is a term found in discussions of systems analysis and time-series modelling.

To give a more precise and narrower focus, the term 'dynamic model' will here cover methods that have all of the following characteristics: epidemiological macro-models; models that deal with transitions in time; models that move beyond formulating and formalising purely cross-sectional relationships; and models that look beyond analysis of time series to the mechanisms and dynamics behind the series.

Domains of research that are excluded by these criteria, but which may have some claim to the term 'dynamic' in a very broad sense, are: simple charting of prevalence or incidence through time; predictive but non-explanatory models; analyses that model the levels of drug prevalence or incidence without reference to explanators, for example very simple auto-regressive or moving average (or Box-Jenkins) approaches; micro-

models that study the progression of individual drug users through various stages of their drug careers, and the process of becoming addicted; micro-models that look at the psycho-dynamics within individuals and their motivations and structures in relation to changes in their drug-taking behaviours; and transmission models of the spread of infection amongst drug users

Issues concerning dynamic modelling

A central reference worth consulting for an overview of dynamic models of prevalence, which includes comparisons with static prevalence-estimation techniques, is the papers contained within the special issue of the *Journal of Drug Issues* on prevalence estimation (Hser and Anglin, 1993). There are, however, several issues concerning dynamic modelling that are worth mentioning specifically.

First, on issues relating to measuring population coverage and classification it should be noted that such methods are mostly directly concerned with simply counting people. Prevalence estimation must also consider the time frame and geographic boundaries because drug use is a dynamic process. Non-users move into the actively using population while current users may cease use; or drug users sometimes move in and out of geographic areas.

The types of drug used and levels of use often need to be considered for several reasons. First, multiple-drug use is common and estimates of users therefore need to take into consideration this overlap of individuals and multiple-consumption occasions in order to avoid repeated counts of the same people. Second, because of the associated severe social and health consequences, some prevalence estimation may focus on a particular type of drug user (e.g. intravenous drug users) or more severe levels of use (e.g. addicts). However, when prevalence estimation is approached from a view of total drug consumption, all categories of drug users must be considered since all levels of drug use contribute. Definitions are necessary to identify appropriate data sources and to provide valid prevalence estimates.

The second point relates to the use of data from national surveys to estimate prevalence. Several national surveys and special-purpose federal data systems containing drug-related information have been conventionally used for prevalence estimation. Major surveys in the United States include the National Household Survey on Drug Abuse, the High School Senior Survey, the National Ambulatory Medical Care Survey and the National Hospital Discharge Survey. Examples of federal data systems include the Drug Abuse Warning Network, the System to Retrieve Drug Evidence, the Uniform Crime Reports, the Drug Use Forecasting Program and the Client-Oriented Data Acquisition Process. These data sources have both strengths and weaknesses and several issues must be considered in assessing the utility of a particular data system for prevalence-estimation purposes. In particular, all of the above data sets are American, and this may reduce generalisability of the published results for EU countries. However, recent initiatives by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) may provide alternative European data sets which are suitable for these types of analyses.

Particular care should be taken over the issue of sampling. Most large-scale surveys of drug use are based on probability-sampling techniques. However, general surveys often undersample certain lower-risk groups because of non-response or non-coverage. If drug users are disproportionately over-represented among the sample population used in the survey then this will overestimate the prevalence of drug use.

It is also worth noting whether the data-storage system is event-based or person-based. Event-based systems are those in which each record arises from a single event, such as emergency-room admission. Person-based systems are those that provide records corresponding to individuals or allow ways to link an individual's multiple records. Except for survey data, most federal monitoring systems are event-based recording systems. It is often the case that several records belong to one individual who has multiple treatment admissions or emergency-room episodes. Thus, the number of people actually responsible for the number of records in these data needs to be determined if the person-based prevalence estimate is desired.

In addition to the availability of data, the description of trends in use over time also requires consistency in the reporting panel, a goal which is not often achieved in continuously reporting indicator systems over long periods of time, especially when reporting by the contributing agencies is voluntary. Inconsistencies in reporting standards and practices must also be considered. The various data systems may also not share standardised methods of data collection, so that the comparison of indicators and their interpretation must be carefully attended.

Finally it should be well understood that prevalence-estimation methods differ in their data requirements and statistical properties and thus in their utility for providing certain types of prevalence estimates. Each of these methods has strengths and limitations, and none can provide estimates without knowledgeable and careful application. Some of these limitations are due to the necessary simplicity of the assumptions of the model, others are due to the demands for specific data of a certain quality. Therefore, in selecting a model, the user must consider the appropriate use of the data that are available and interpret estimates derived from these methods within the appropriate context.

Types of prevalence-estimation methods

The comments in this chapter describe several prevalence-estimation methods with either historical importance or promising future applications. These methods include: synthetic estimation (population-projection models and principal-component approaches); closed-population capture models (specifically the truncated Poisson estimation model); open-population models (specifically the ecological models and the Markov-based dynamic-recapture models); and system-dynamic modelling.

The useful distinction among these methods is to categorise them as primarily dynamic versus static approaches. Static methods such as synthetic estimation or closed-population models evaluate a population at a single point in time. Dynamic methods such as open-population models or system-dynamic models are appropriate when a population is traced over time. The choice of model is largely dictated by the data that are available.

To highlight the contrast between these two broad approaches some brief description of the static approaches is appropriate. These are the least complex of the prevalence-estimation models, describing the system at a single point in time. Because of their convenience, these models have been the most popular in drug-use-prevalence estimation.

Synthetic estimation is the simplest of the static-estimation techniques and is the main category of static models that is used in practice. These methods develop prevalence estimates for new populations using readily available data sources or indicators from known populations by matching various predictor variables, usually demographic characteristics, and determining appropriate weighting schemes. The following two methods are used to determine the proper weights to be applied.

Projection models are based on the logic that if the drug-use prevalence rates are known in a population with a known demographic distribution, then the relationships between prevalence and demographic characteristics can be transferred or projected to another population, either smaller or larger than the first.

Principal-component-based approaches assess the relationships observed among multiple indicators in several geographic areas in an attempt to obtain a single composite and common indicator of drug use. By combining several indicators into a single composite index with the weights determined by the principal-components analysis, an index that is more reliable than any single indicator alone may be derived.

Dynamic models differ from the static models in that they describe processes over time. The process examined may be the states of individual drug users, as in the open-population multiple-capture models, or it may be a collection of aggregate societal states, as in system-dynamic models. Because these models can represent processes more accurately than static models, and because they typically incorporate more information, these models potentially result in more accurate estimates of prevalence.

Synthetic-estimation techniques are clearly not dynamic and will not be discussed further. Yet, although closed-population capture models are strictly static models, they are of historical interest and will be considered further in the next section. The following sections will then respectively cover the open-population capture models and the system-dynamic models.

Closed-population capture models

A major heading for prevalence estimation is closed-population multiple-capture methods. These techniques are not usually framed as dynamic models, in that they estimate a population at a single point in time, or at least a closed population assumed closed over a fixed time interval; as such they are usually categorised as a variety of static estimation. In these techniques, two or more surveys of different data sources, such as emergency-room and arrest records, are used to probe the given population during a single time frame. Each survey must be able to identify specific cases and determine which individuals have been detected in both of the surveys. Using the

information about the relative sizes of the samples and their overlaps, the number of individuals that have not been detected can be estimated. This procedure has also been referred to as dual-system estimation (Chandra-Sekar and Deming, 1949; Ericssen and Kadane, 1985).

The population whose size is to be estimated here is said to be closed because the single sampling time implies necessarily that it cannot change and gives the appearance of a simple cross-sectional study. It is, however, likely that the capture points cannot all be considered as generating a sample simultaneously with unfettered potential for overlaps between these samples. In reality, a person appearing at one capture point can only appear at another after a time lag of a length that is dependent on the precise nature, physical or administrative, of the capture points. This consideration throws the multiple-capture process immediately into a dynamic framework even when the population being estimated is closed to change over a time period. It would appear, however, that no studies or methods have specifically focused on this structured capture process to date. This topic is not pursued further here.

The truncated Poisson model

The truncated Poisson estimation model is the simplest version of the multiple-observation models. It can be applied in situations where only the frequency with which an individual appears in a data system is recorded. In this procedure, events, such as arrests for drug possession, that occur one or more times to individuals in a population during a specific interval of time are examined. The distribution of people experiencing various numbers of events is used to extrapolate the size of the population. In the population whose size is to be estimated, some individuals have no events and are never observed, some have one event, some have two, and so forth. The problem is that although the number of people with one, two, three or more events can be observed, the number with no events cannot be directly observed. However, by assuming a probability model, this can be used to extrapolate what the number of people with no events is, and hence estimate the total population.

Now when the events are rare, as they usually are for drug-use incidents, then under the following two assumptions it can be deduced that a plausible distribution for the number of events is a Poisson form, leading to truncated Poisson estimates (Blumenthal et al., 1978). The two assumptions about the events needed for it to be a Poisson distribution are: that they occur randomly over both individuals and times; and that their rate is homogeneous across the population. To estimate the size of the population, an incomplete Poisson distribution is fitted to the frequencies of the observed events, and the single-rate parameter of the distribution is estimated. Knowing this parameter, the size of the unobserved category is estimated and added to the observed count to obtain the final estimate. Once the data have been fitted to the Poisson distribution, the observed and expected frequencies for each number of events can then be compared via a chi-squared test, which will facilitate validation of these assumptions, or not, as the case may be.

The truncated Poisson models have been used for estimating the size of the criminal population from arrest history records (Greene and Stollmack, 1981) and the number of persons engaged in using or selling drugs from drug-related arrest data (Woodward et al., 1987). In these applications, an arrest distribution was constructed from the number of observed arrests and the number of arrestees responsible for them, then the truncated Poisson estimation procedure was applied to derive the population estimates. Based on a similar rationale, the Research Triangle Institute (1988) applied the model to estimate the size of the treatment-susceptible heroin population. Their model was unusual in that two separate sources of data were used to estimate the Poisson rate. The number of treatment admissions was available from one data source, while the distributional information was derived from a separate, non-linked source. In this two-source implementation, the comparability of the two populations is critical to ensure that the assumed Poisson distribution is applicable.

The strength of the truncated Poisson method lies in the simplicity of its data requirements and its straightforward statistical formulation. As long as the data can be consolidated into a frequency distribution of the number of people at each level of the repeated observation, an estimate is easily obtained. However, the quality of the estimates depends on the degree to which the Poisson model is an adequate description of the underlying distribution. In particular, the counts must be independent Poisson events. This assumption is frequently violated. For example, criminals are strongly motivated to avoid re-arrest and are, to some extent, quite successful in doing so. On the other hand, risk of arrest may increase as the offender becomes known to the police. The effects of such violations of independence on the truncated Poisson-derived estimates are often not calculable.

Open-population capture models

Data that are drawn from identifiable individuals in a system of successive surveys or censuses using the open-population multiple-capture method are similar to those obtained in the closed-population multiple-capture situation described above. These data also form an incomplete contingency table. The similarity is superficial, however, for the processes to which the models are applicable are quite different. In the closed-population situation, every individual is at risk for every census, while in most longitudinal repeated censuses, some individuals leave the population before the final census while others enter after the first census is complete. Therefore, closed-population models trace and estimate the size of a single population, while open-population models keep track of population changes and provide estimates reflecting such time-related changes of in and out flows. The data from a longitudinal dynamic process thus require a different type of statistical model, based on some assumptions about the open population under consideration.

Most applications of the multiple-capture methodology to longitudinally repeated drug-use samples have used closed-population models (e.g. Doscher and Woodward, 1983; French, 1977; Greenwood, 1971; Woodward et al., 1985). Such estimates are potentially biased because some individuals may not be available throughout the entire sequence

of time sampling and the degree of bias may increase with longer intervals between samplings. The nature and magnitude of these biases have not been studied. However, the magnitude of the standard errors in these closed-population applications is usually quite large, ranging from 10 % to 80 % of the estimated population sizes. Such poor estimation is especially serious within geographical areas representing relatively small populations. Models that accommodate the more realistic open-population dynamics may be generally more appropriate for estimation purposes.

A more sophisticated approach using repeated sampling that builds on the truncated Poisson model described above may provide a better estimation methodology. Ecological open-population models, which assess the size and character of a biological population based on repeated marked samples, have been developed and extensively analysed (for reviews see Seber, 1982, 1986). The most common class of ecological sampling models that are applicable to the estimation of the number of drug users is the capture–recapture type. For an open population, the most recent models are the Jolly–Seber model (Jolly, 1965, 1982; Seber, 1965, 1982) and a related model by Cormack (1981, 1985). However, a number of the assumptions of these models are unrealistic for drug-using populations, and none of these models has yet been applied to drug-use prevalence estimation. For example, the Jolly–Seber model assumes that: all samples are of independent and identical capture probability and survival probability; samples are instantaneous and release is made immediately after each sampling; the drug user's behaviour is unaffected by the capture history; and there is no temporary emigration from the population. These assumptions can hardly be satisfied by existing data or by the population characteristics of drug abusers.

Markov dynamic-recapture models

Another variety of the open-population multiple-capture approach developed recently is a Markov-based dynamic-recapture model. Instead of counting the individuals captured in each sample, a longitudinal model is based on the variety of capture histories. The model characterises capture probabilities by a two-step sampling process. The initial sampling probability is governed by a stochastic process in which users are drawn from a large population of non-users. After this first observation, the balance of the process is governed by the dynamics of a state structure that represents the evolution of drug-consumption patterns and their repeated observations by some indicator system (e.g. treatment admissions). This process forms a Markov chain (Wickens, 1993). The full history probabilities are therefore the product of three terms: the size of the sample in which an individual is first observed; the probability of that observation; and the probability of the observation history subsequent to the initial observation, characterised by the full variety of such histories. This procedure generates estimates of the size of the population from which the observations are drawn.

In this procedure, the data from a repeatedly observed population are treated as sampling histories of individuals. As in the closed-population multiple-capture models, a probabilistic model is fitted to the observed histories, then extrapolated to count the unobserved members of the population. The Markov and semi-Markov models describe

the way the individuals change between observations as transitions among a small set of discrete states. Individuals enter into the target population, pass through various states of observation and non-observation, and possibly may leave the target population. In a semi-Markov process, the individual's passage among the states is modelled in continuous time by a homogeneous transition process. For data from a series of samples, the transitions are only observed at the times of sampling, and a simpler discrete-step Markov chain is used.

The essential feature of a Markovian process is the assumption that the current state of an individual provides all available information about his or her future behaviour. How the individual initially reached that state is irrelevant. This independence of history means that the properties of the process are summarised by a set of transition rules pertaining to the specific states. For example, individuals in the observed state are characterised by the three probabilities (or rates) that indicate how likely they are to pass to each of the three states to which the observed state is connected.

The transition probabilities among the observation and post-observation states can be estimated from the observed histories and are used to infer the target population size. The crucial assumption underlying these methods is that the dynamics of unseen individuals match those of the individuals who have been observed. Moreover, the transition structure must adequately approximate the dynamics of real individuals. It should be recognised that these models impose a considerable structure on the data and cannot be accurately used without a history sequence of some length.

The Markov estimation technique has the advantage of being able to provide a somewhat more realistic dynamic description of the drug-using process. Its weakness is that the model comes with some strong intrinsic statistical assumptions of homogeneity and independence of history. In addition, to be practically applied, the model can have only a minimal dynamic structure. The limitation on the complexity of the model is necessary if its parameters are to be identifiable and estimated. A rich data source is needed to identify any complex structure, and such comprehensive data are not usually available.

Although the open-population models potentially allow a more realistic picture of drug-using populations as they evolve over time than do closed-population models, such models still require certain restrictive assumptions. Some of these assumptions are particular to the specific models developed in ecological situations, and others serve to simplify the required statistical models so that parameters can be adequately estimated. The latter assumptions include the requirements that individuals behave independently of each other and that the model's parameters are homogeneous throughout the population. A number of applications have shown that these assumptions are often violated in biology and health sciences (e.g. Huber, 1962; Manly, 1971; Wittes, 1974), and simulation studies have shown potentially large biases in population estimates under such conditions (e.g. Carothers, 1973; Gilbert, 1973).

Theoretically, surveys or other data sources used by multiple-capture methods (either closed or open) should be comprehensive in population coverage. In reality, data suitable for this type of application have been limited to treatment-admission records that do

not provide such comprehensive coverage, thus limiting the generalisability of the estimation results. One common difficulty in applying multiple-capture models is the necessity of matching individuals across observations. The data source must ensure that an individual captured in one sample is identifiable as the same person if captured in another sample. However, such information is usually difficult to obtain.

Because the multiple-recapture models attempt only to estimate the number of unobserved individuals from the observed sample, they cannot extrapolate to completely unobserved sub-populations. Like any of the statistical methods, they cannot solve the hidden-population problem. However, open-population models generally are based on probability theory and have better-known statistical properties that allow estimates to be evaluated by standard methods such as confidence intervals. In addition, boundaries of generalisability can usually be inferred by data coverage and model specifications.

System-dynamic modelling

System dynamics is a general methodology, first developed in the late 1950s, for analysing dynamic phenomena through the use of simulation models based on information-feedback-control theory. A system-dynamic model consists of an interconnected set of difference equations representing continuous time movement and accumulations of people, materials and information. After being assigned initial conditions consistent with historical data, this set of equations is used to generate output over time. If the model is a valid one, this output will closely mimic the true course of events, and the model may be used for prevalence estimation and for making conditional forecasts.

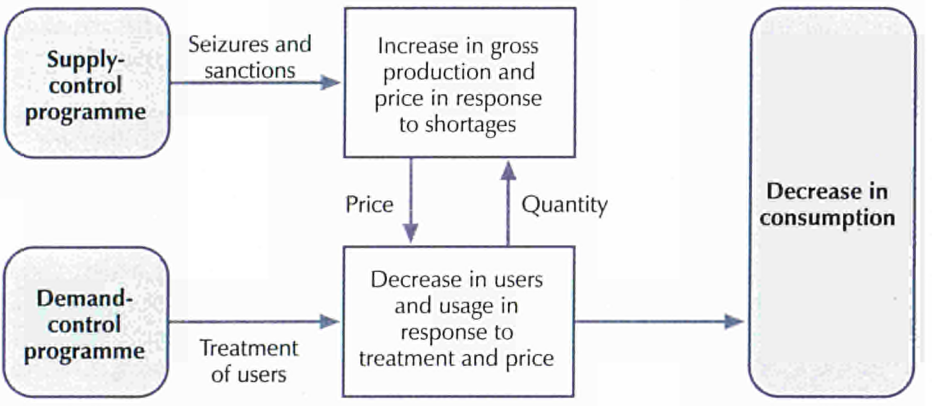
An example of system-dynamic modelling studies of illicit drug use include the 'Persistent Poppy' model of Levin et al. (1975). The Persistent Poppy model examined heroin use in New York City from the standpoint of policy rather than prevalence estimation. Although the model contains several interesting endogenous factors, such as law-enforcement activity and treatment programmes, it was developed at a time when the numerical data needed for its calibration and validation were lacking.

Another example of a system-dynamic model was developed by Gardiner and Shreckengost (1985, 1987; Shreckengost, 1984, 1985). The models developed by Gardiner and Shreckengost address drug supply and demand on a national level, with specific application to heroin and cocaine. They have been used primarily to make inferences about drug-import levels, thus focusing on quantity measures rather than persons or events. The model developed for heroin was adapted to estimate the number of users, but such estimates were shown to be rather sensitive to uncertain assumptions about the number and type of drug-user categories (Shreckengost, 1984). The models of Gardiner and Shreckengost may also lack sufficient feedback structure and internally generated momentum to be useful for prevalence estimation and forecasting.

A recent model of cocaine use (Homer, 1990) was developed to make inferences about unobserved populations from available data sources related to drug-use prevalence. This model reproduces historical drug-indicator data from 1976 onwards and produces prevalence estimates and forecasts for several population categories, including casual and compulsive users. For this purpose, the model uses two sorts of information. First, it uses information about the logical relationships among various population categories. For example, the model depicts the escalation process by which casual users become compulsive users, and distinguishes the effects of powder cocaine from those of crack. Second, the model uses information about the logical relationships between the population categories and other indicators, including morbidity and mortality, drug-crime arrests, retail price and purity, and so forth. For example, the model is calibrated to reflect the idea that compulsive users are more likely than casual users to show up in the Drug Abuse Warning Network (DAWN) data on cocaine-related morbidity and mortality. With sufficient numerical data and knowledge of a phenomenon's structure and dynamics, the range of estimates for the hidden populations may be narrowed to a considerable degree. This model has been further used by Everingham and Rydell (1994) and Rydell and Everingham (1994) to model the demand for cocaine and to explore various scenarios developed from current drug policies (see Figure 2).

More technically in the system-dynamic technique, quantities such as the prevalence of drug use are measured by embedding them in a larger model that describes the feedback relationships among these quantities and other social and economic factors. A system-dynamic model describes continuous-time change in quantities and factors through an interconnected set of differential equations. System states, such as the number of individuals in a particular state of drug use, are represented by level variables L(i). The level variables are related to one another and change according to net-rate functions

Figure 2: Logical structure of cocaine-control model



Source: Rydell and Everingham (1994)

f; for each level L there is a single net-rate function f_L. These net-rate functions use intermediate variables known as rates and auxiliaries. Rate variables define identifiable flows into and out of the levels, such as drug-use initiation, escalation, stopping and relapse. Auxiliary variables define other concepts that make the model more intuitive and natural, such as retail price and perceived health risk. The net-rate functions are parameterised by constants C, which include delay times and strength-of-response parameters, such as price elasticities. The net-rate functions may also be influenced by exogenous variables X(t) that change over time and are determined by external data, such as the baseline population growth rate and the fraction of drug imports that are seized. With the many level and exogenous variables, constants and net-rate functions written as vectors, a system-dynamic model is expressed as

$$dL(t)/\Delta t - f_L(t) = f[L(i), X(t); C]$$

When investigating the model, these differential equations are typically not solved analytically, but are expressed as difference equations, for small time intervals Δt.

$$L(t) = L(t-\Delta t) + \Delta t \times f[L(i), X(t-\Delta t); C)$$

The variables are simulated starting at some initial time such as $t = 1976$ and stepping forward to subsequent times. The computation interval Δ has no real-world significance and is chosen to be sufficiently small such that the simulation output lies close to the true solution of the corresponding differential equations.

The validation of a system-dynamic model, unlike that of the simpler probabilistic models described elsewhere, is an ongoing process of building confidence in the realism of the model's structure and behaviour, using a variety of largely qualitative tests (Forrester and Senge, 1980). The complexity of the models allows for similar observed behaviour to be generated by a variety of structures, which must be carefully chosen if unobserved dynamic quantities, such as prevalence, are to be accurately described. Confidence in the model and its estimates is enhanced when all equations have concrete real-life significance, are dimensionally correct and operate appropriately even under extreme conditions. Confidence is also enhanced when the model faithfully recreates the dynamic patterns and correlations observed in real life, and when it bring to light behaviour in the real system that has gone unrecognised or unexplained (Mass, 1991).

System-dynamic models typically attempt to explain observed dynamics as being the consequence of endogenous feedback relationships among constituent variables. This endogenous perspective distinguishes the system-dynamic approach from other modelling methods, such as synthetic estimation, which depend heavily upon exogenous or independent predictors whose own behaviour over time is left unexplained by the model. The continuous-feedback perspective of system dynamics also leads to models that contain a greater variety of system variables than the multiple-capture or Markov-type models, and tend to be larger in scope. This enlarged scope is rendered manageable by modelling flows as aggregate measures, rather than by keeping a unique record of every individual unit in the flow, as is done in the multiple-capture models.

System dynamics is an attractive approach for prevalence estimation largely because it can be used to explain past history, fill in the gaps in indicator data, and project outcomes under different assumed scenarios and policy interventions. In addition, a system-dynamic model can often help detect certain flaws in existing indicator data sets, such as possible logical inconsistencies between incidence data and prevalence data. However, the very flexibility of system dynamics opens the door to potential model mis-specification, a danger that becomes greater as the number of variables and conceptual scope of the model increase relative to the quantity of relevant data. Also, like any method, the accuracy of a system-dynamic model is sensitive to the quality of the data used to calibrate it. Nevertheless, a well-specified system-dynamic model may be useful even if it falls short in numerical precision, because it reveals and anticipates trends that other methods may miss.

Because of the complexity of system-dynamic models and the risks of mis-specification, a variety of tests for building confidence into such models have been offered that go well beyond the usual requirement that historical data be reproduced (Forrester and Senge, 1980). But these validation techniques are themselves subject to uneven application or improper interpretation. It must be recognised that system-dynamic modelling, despite its many attractions, is difficult to master, and there are pitfalls in its application that must be carefully avoided.

Conclusions

Techniques are already available for the dynamic system analysis of drug prevalence. The problem with such models is that their applicability is highly dependent on the level of data available. As better-quality and more comprehensive data become available it may be that more complex modelling techniques will also be proposed; however, at present there are already methods proposed which go beyond the available data. As such, the key priority in this area must be the implementation of better and more comprehensive methods of data collection. In addition, if these techniques are expected to be applied to the European drug market then it is also important that European data sources are developed.

References

Blumenthal, S., Dahiya, R. D., Gross A. J. (1978). 'Estimating the complete sample size from an incomplete Poisson sample'. *Journal of the American Statistical Association* 73: 182–7.

Carothers, A. D. (1973). 'Capture–recapture methods applied to a population with known parameters'. *Journal of Animal Ecology* 42: 125–46.

Chandra-Sekar, C., Deming, W. E. (1949). 'On a method of estimating birth and death rates and the extent of registration'. *Journal of the American Statistical Association* 44: 101–15.

Cormack, R. M. (1985). 'Examples of the use of GLIM to analyze capture–recapture studies', in Morgan, B. J. T., North, P. M. (eds) *Statistics in ornithology*, Lecture notes in statistics, 29. Berlin-Heidelberg: Springer-Verlag.

Cormack, R. M. (1981). 'Loglinear models for capture–recapture experiment on open populations', in Hiorns, R. W., Cooke, D. (eds) *The mathematical theory of the dynamics of biological populations*. London: Academic Press.

Doscher, M. L., Woodward, J. A. (1983). 'Estimating the size of subpopulations of heroin users: applications of log-linear models to capture–recapture sampling'. *The International Journal of the Addictions* 18: 167–82.

Ericssen, E. P., Kadane, J. B. (1985). 'Estimating the population in a census year: 1980 and beyond'. *Journal of the American Statistical Association* 80: 98–109.

Everingham, S. S., Rydell, C. P. (1994). *Modelling the demand for cocaine*. Santa Monica, CA: Drug Policy Research Centre, RAND.

Forrester, J. W., Senge, P. M. (1980). 'Tests for building confidence in system-dynamic models'. *TIMS Studies in Management Sciences* 14: 209–28.

French, J. F. (1977). 'Prevalence of heroin use in New Jersey'. Unpublished report of the Department of Health, Alcohol, Narcotic, and Drug Abuse Unit, State of New Jersey.

Gardiner, L. K., Shreckengost, R. C. (1987). 'A system dynamics model for estimating heroin imports into the United States'. *System Dynamics Review* 3 (1): 8–27.

Gardiner, L. K., Shreckengost, R. C. (1985). 'Estimating heroin imports into the United States', in Rouse, B. A., Kozel, N. J., Richards, L. G. (eds) *Self-report methods of estimating drug use: current challenges to validity*, NIDA Research Monograph 57, DHHS Publication No. (ADM)85-1402. Washington DC: US Government Printing Office.

Gilbert, R. O. (1973). 'Approximations of the bias in the Jolly–Seber capture–recapture model'. *Biometrics* 29: 501–26.

Greene, M. A., Stollmack, S. (1981). 'Estimating the number of criminals', in Fox, J. A. (ed.) *Models in quantitative criminology*. New York: Academic Press, 1–24.

Greenwood, J. A. (1971). 'Estimating the number of narcotic addicts'. Drug Control Division, Office of Scientific Support, Bureau of Narcotics and Dangerous Drugs, US Department of Justice Document No SCD-TR-3.

Homer, J. (1990). 'A system dynamics simulation model of cocaine prevalence'. Final report to the National Institute of Justice, Grant 87-U-CX-0042.

Hser, Y.-I., Anglin, M. D. (1993). 'Prevalence estimation techniques for drug-using populations'. *Journal of Drug Issues* 23 (2): 163–360.

Huber, J. J. (1962). 'Trap response of confined cottontail populations'. *Journal of Wildlife Management* 26: 177–85.

Jolly, G. M. (1982). 'Mark-recapture models with parameters constant in time'. *Biometrics* 38: 301–21.

Jolly, G. M. (1965). 'Explicit estimates from capture–recapture data with both death and immigration: a stochastic model'. *Biometrika* 52: 225–47.

Levin, G., Roberts, E. B., Hirsch, G. B. (1975). *The persistent poppy*. Cambridge, MA: Ballinger.

Manly, B. F. J. (1971). 'Estimates of a marking effect with capture–recapture sampling'. *Journal of Applied Ecology* 8: 181–9.

Mass, N. J. (1991). 'Diagnosing surprise model behavior: a tool for evolving behavioral and policy insights'. *System Dynamics Review* 14: 209–28.

Research Triangle Institute (1988). Presentation at the Prevalence Estimation Technique Conference, Washington DC, November.

Rydell, C. P., Everingham, S. S. (1994). *Controlling cocaine: supply verses demand programs*. Santa Monica, CA: Drug Policy Research Centre, RAND.

Seber, G. A. F. (1986). 'A review of estimating animal abundance'. *Biometrics* 42: 267–92.

Seber, G. A. F. (1982). *The estimation of animal abundance and related parameters*. London: Macmillan.

Seber, G. A. F. (1965). 'A note on the multiple-recapture census'. *Biometrika* 52: 249–59.

Shreckengost, R. C. (1985). 'Estimating cocaine imports into the United States'. Unpublished manuscript.

Shreckengost, R. C. (1984). 'How many heroin users are there?' Unpublished manuscript.

Wickens, T. D. (1993). 'Quantitative methods for estimating the size of a drug-using population'. *Journal of Drug Issues* 23: 185–216.

Wittes, J. T. (1974). 'Application of a multinomial capture–recapture model to epidemiological data'. *Journal of the American Statistical Association* 69: 93–7.

Woodward, J. A., Bonett, D. G., Brecht, M. L. (1985). 'Estimating the size of a heroin-abusing population using multiple–recapture census', in Rouse, Kozel, Richards (eds) *Self-report methods of estimating drug use*, 58–171.

Woodward, J. A., Brecht, M. L., Bonett, D. G. (1987). 'Statistical analysis of drug abuse indicators'. Final report to the Drug Enforcement Administration and the National Institute on Drug Abuse (NIDA), Washington, DC.

CHAPTER 9

STRUCTURAL EQUATIONS AND PATH ANALYSES: USING MULTIPLE INDICATORS OF DRUG PREVALENCE

Colin Taylor and Philip Young

In many areas of social research it is sometimes not possible to measure directly the concepts that are of major interest. The first distinction that researchers wish to make is that between *frequency* of drug-taking and the *quantity* of drugs taken. Distinguishing between frequency and quantity has important implications for recognising the presence of drug or alcohol problems by health-care professionals, employers or co-workers. Many studies combine frequency and quantity measures to assess drug and alcohol consumption (e.g. Makela and Simpura, 1985).

However, frequency of drug use and quantity or amount of substance ingested per occasion may be differentially related to *disruptive* or 'problem' use of drugs. Thus, although frequency and quantity are typically correlated in a positive direction (Fitzgerald and Mulford, 1984; Huba et al., 1982), they may not be equally complicit in predicting general problems with drug use or use of drugs in inappropriate settings (i.e. disruptive use at work or school). Huba et al. (1982), in a study of 257 adolescents between 15 and 18 years of age, found a correlation of 0.91 between quantity and frequency measures of use of alcohol, cannabis and hard drugs. In addition, researchers are often interested in the public's subjective perception of the drug problem.

From the above, four underlying concepts related to drug use that may interest researchers can be identified: the *frequency* of drug use; the *quantity* of drug use; the amount of *disruptive* drug use; and the subjective *perception* of drug-use problems. Researchers may then be interested in the relationships between these concepts or factors. Specifically, researchers are often interested in modelling the correlations between them. The idea of using these underlying factors first appeared in the field of psychometric testing when researchers came to realise that the measurement of hypothesised constructs, such as personality traits and intelligence, could be improved by considering scores on a variety of indicators (Bentler, 1990).

Unfortunately, although it is easy to discuss drug usage conceptually in terms of these factors, it is often difficult to measure these quantities directly. As a result, researchers often tend to collect information on variables which they believe will act as indicators of the underlying concepts. The most basic way of proceeding is to use a technique known as factor analysis. The basic factor-analysis model is essentially the same as

multiple regression, except that the observed multiple indicators are regressed on the unobserved underlying factors, which are often referred to as latent variables.

When a specific number of factors are assumed to be involved, then the factor analysis really revolves around trying to justify, or confirm, that these are the correct factors to use, and is often called confirmatory-factor analysis. The use of multiple indicators has long been recommended as an alternative to counting actual prevalence or incidence, and at present the Pompidou Group of the Council of Europe is still making their use a central part of its recommended programme. The use of a multi-trait multi-method procedure (Campbell and Fiske, 1959) using latent-variable confirmatory-factor analyses is commonplace.

The confirmatory-factor model has been described by Schmitt and Stults (1986) as the preferred method of analysing multi-trait multi-method matrices. Such procedures have been successfully used to demonstrate construct validity and to derive more accurate measures of drug use (Stacy et al., 1985), while Harvey et al. (1985) found it useful to incorporate method factors into their confirmatory-factor analysis of the job diagnostic survey.

The above examples of latent variables are, of course, only some of the underlying constructs, or factors, that have been postulated. Another important aspect to recognise in researching drug prevalence is that poly-drug abuse is very common. In fact, most drug users do not limit themselves to a single substance and typically use two or more drugs as part of their lifestyle (Clayton and Ritter, 1985). In particular, those who use illicit drugs (i.e. cannabis, cocaine or other hard drugs) also tend to use alcohol (see Newcomb and Bentler, 1986a, 1986b). Thus, the general drug factors identified in these analyses reflect poly-drug use, a more realistic appraisal of drug involvement than studying single substances.

However, most heavy users of drugs other than alcohol tend to be poly-drug users (Clayton and Ritter, 1985), with alcohol being one of the multiple substances. Alcohol use in a poly-drug-use context may have quite different consequences than when it is used in moderation. It is possible that moderate use of drugs other than alcohol may not invariably be associated with problem use of such substances by young adults.

Another area where drug-prevalence research is concentrated is in the detailed relationship of drug use to personal behaviours and in particular to drug-related problems, expressed as underlying latent constructs. It is of particular interest to determine whether high levels of drug use are invariably associated with inappropriate or disruptive use of drugs on the job, or at school, or with other specific problems of use (e.g. car accidents and arrests) among adolescents. In the same way, many features of intravenous drug users' (IDUs) behaviour, such as sharing injecting equipment or sexual activity, may be considered latent variables with many partially correlated indicators.

Sadava and Secord (1984) reported that consumption of alcohol and drugs was related to, but not equivalent to or redundant with, alcohol problems. Sadava (1985) also found that self-reported consumption and problems were largely independent among

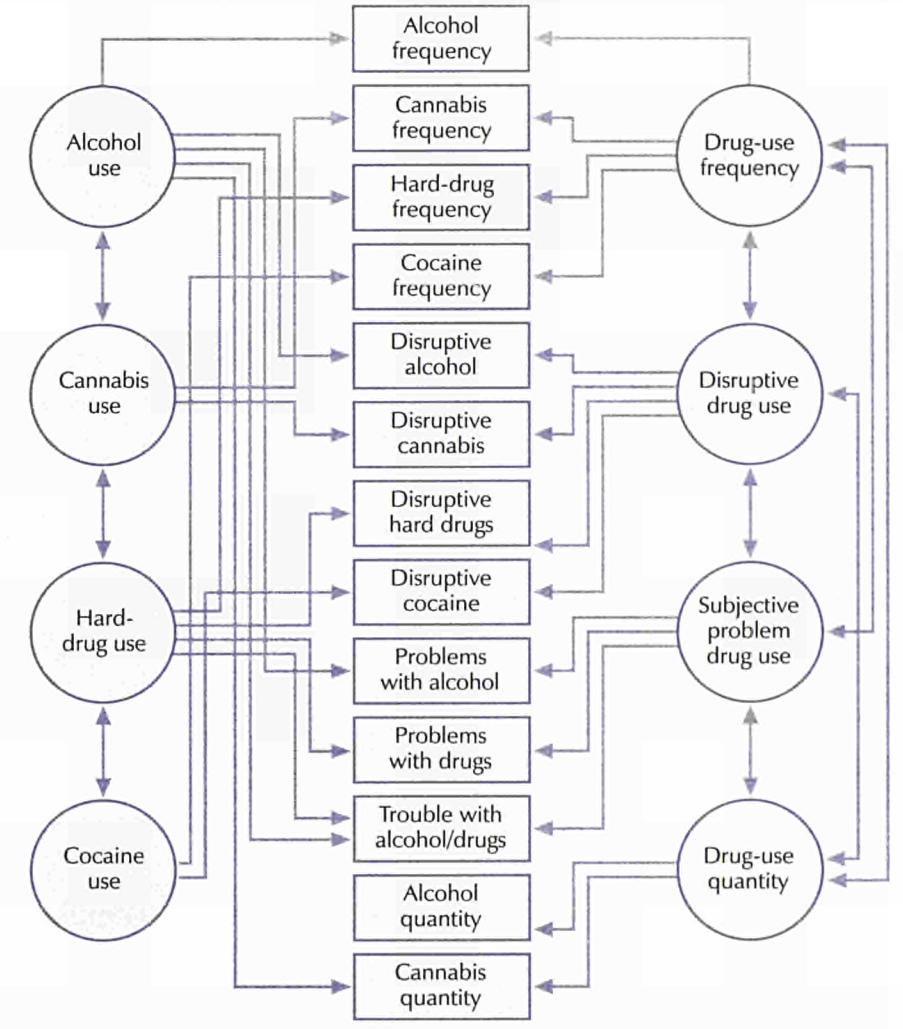
adults. He argued that alcohol consumption and alcohol-related problems must be more clearly separated on conceptual and methodological bases. Survey studies typically have found an average correlation in the 0.3 to 0.4 range between measures of consumption and problems (Sadava and Secord, 1984). In a longitudinal study, Stein et al. (1987) found different prior determinants of use and problem use of drugs in young adulthood. General use was related to prior use and early adult models, whereas problem use was related to earlier rebelliousness.

In their 1988 study, which is a good example of the genre and which touches on all three of these issues, Stein et al. examined the interrelations among different ways to assess drug use, self-reported problems with use and specific negative outcomes resulting from drug use. Specifically, the analyses concern the relation among self-reported use of drugs in the workplace; arrests for driving while intoxicated (DWI) and for other drug-law violations; having car accidents while drunk or stoned; general drug problems; and frequency and quantity of using alcohol, cannabis, cocaine and other hard drugs.

Figure 1 depicts the hypothesised confirmatory-factor model and illustrates how the specific drug-use components (alcohol, cannabis, cocaine and hard drugs) are separated from poly-drug-use tendencies (frequency and quantity), disruptive use and subjective problem-use factors. These constructs are identified by 13 measured or observed variables. Latent variables are depicted as circles, and manifest variables are represented as rectangles. On the left-hand side of Figure 1, a general alcohol-use latent variable is hypothesised to generate measures of alcohol frequency, alcohol quantity, disruptive alcohol use, subjective problems with alcohol, and trouble with alcohol or drugs. Similarly, the other measured variables of drug use are assumed to reflect specific tendencies to use cannabis, hard drugs and cocaine. On the right-hand side of Figure 1, a substantive general latent variable of drug-use frequency (i.e. frequency of poly-drug use) is assumed to generate manifest measures of frequency of alcohol, cannabis, hard drug and cocaine use. Drug-use quantity is identified by quantity measures of alcohol and cannabis use. The tendency to use a range of drugs on the job or in school is called disruptive drug use and is reflected in four measures of disruptive alcohol, cannabis, cocaine and hard-drug use. Three measures of subjectively perceived problems with drugs or alcohol, both long-term and more recent, are reflected in a latent variable of subjective problem drug use. The lines with double-headed arrows indicate co-variances allowed between the latent variables. Correlations between the substantive latent variables and the specific drug-use latent variables were set at zero.

The report examined relations among self-report measures and outcomes of drug use among 739 young adults. It is worth noting that the usual factor-analysis model assumes normality of the indicators. However, the indicators used by Stein et al. (1988) were screened for appropriate distributional properties, such as skewness and kurtosis. Most of the indicators were found to have distributions quite far from normality. Fortunately, confirmatory-factor analysis is known to be fairly robust to departures from normality.

Figure 1: Confirmatory-factor model of multi-trait multi-method assessment of drug use



Source: Stein et al. (1994)

Predicting single traits from multiple indicators

Confirmatory-factor analysis considers a number of latent variables as the underlying factors that explain the correlations between the observed multiple indicators, so that if the value of the latent variables were known there would be no partial correlations between the indicators. Although such a model is useful in understanding the associations between the various factors and indicators, it is not intended to provide predictive estimates.

One advance would be to form a predictive equation for one of the observed indicators, for example overall drug use, and predict this using the other indicators through the latent variables. One major line of development in this area has been the study of geographic variation in prevalence levels, derived from national, or rather sub-national, official data sources. Multiple-indicator methods that use a combined multivariate analysis in deriving a single prevalence figure have been used extensively in the United States to chart geographic variation in opiate abuse, particularly at the level of the small metropolitan statistical area (SMSA) in conjunction with the Drug Abuse Warning Network (DAWN) which produces data at this level of aggregation for 24 areas.

Indirect estimation has progressed from simple use of hepatitis indicators (Minichiello, 1974) and national estimates of prevalence (Greenwood and Crider, 1978) to multiple-indicator approaches at both SMSA (Person et al., 1977; Riley et al., 1981) and county level (French, 1977).

Historically, these methods have often gone under the name of 'multi-method single-trait' modelling and more recent methods have grown from the initial models of confirmatory-factor analysis (Joreskog, 1971) developed in psychological research methodology. Additionally, a variety of demographic models of inter-SMSA variation in indicator-derived prevalence (Schlenger and Greenberg, 1978) have been employed to investigate inter-metropolitan correlates of drug use, and demographic models of census-tract variation in drug-abuse prevalence have been tested by Demaree et al. (1978).

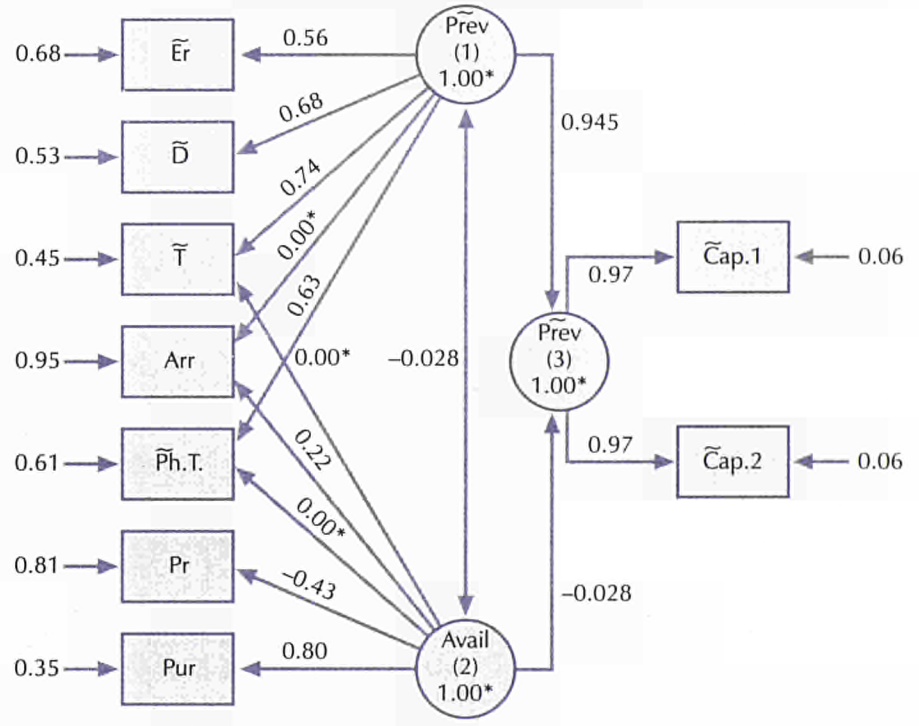
More recently interest has focused specifically on questions of validity and reliability. Beginning with Person et al. (1977), the conceptual problems of estimating heroin-abuse prevalence from indicators have been discussed. Because of definitional difficulties, they use the term 'heroin abuse' rather than 'heroin prevalence' and consider the use of the term 'heroin problem index'. Earlier papers had assessed the internal consistency of such multi-method single-trait estimates, for example Person et al. (1976) for a set of heroin-abuse indicators drawn from medical, treatment and law-enforcement sources for the 24 DAWN SMSAs.

Indirect support for face validity of the procedure has in the past been demonstrated. Woodward (1978) fitted a single-factor model simultaneously at two time periods separated by a single quarter (three months). Not only could the longitudinal model be accepted as plausible, but the estimated correlation between the heroin-abuse constructs across time was high, indicating the expected high degree of stability across the three-month time period; similar relationships were found by Stoloff et al. (1975) in a longitudinal study carried out within a single SMSA. Moreover, the two-point longitudinal model was fitted in the context of two drug-abuse-related demographic characteristics of the communities – urbanicity and employment. Again, the model showed adequate fit, and the significant coefficients of the model were consistent with the interpretation that the single factor underlying the indicators is a measure of the intensity of heroin abuse: that is, the heroin-abuse construct was positively related to urbanicity and negatively related to employment.

Structural-equation modelling

With the availablity of more powerful analysis programmes (e.g. LISREL, EQS) the trend is now to place such analyses within the currently popular format of structural-equation modelling (SEM). The model used, for example, by Woodward et al. (see Figure 2), can be fitted with a confirmatory-factor-analysis procedure. In that all links between constructs or traits have to be bi-directional and can be regarded as partial correlations, and all links from constructs are unidirectional towards the observed data or methods, the model is restricted to being a confirmatory-factor-analysis model. However, Woodward et al. (1984) used a more versatile structural-equation programme to fit the confirmatory factor-analysis model.

Figure 2: Standardised maximum-likelihood estimates and goodness-of-fit test for the log-rate three-construct model



Notes: X^2 (27) = 43.96
p = 0.03
Er = emergency-room episodes
D = examiner episodes
TR = treatment episodes
ARR = arrest
Ph.T. = pharmacy thefts
Pr = retail price
Pur = retail purity
Prev = prevalence
Avail = availability
Cap = capture

Source: Woodward et al. (1984)

The model used by Woodward et al. (1984) was constructed with multiple traits, the constructs of heroin abuse and heroin availability. With the recognition that the central construct in these indirect indicator models, even if called heroin-use prevalence, is actually a more general and broader concept than simple head-counting and deserving of a different name (Woodward et al., 1984), it becomes useful to have more than one such unobserved construct incorporated into the analysis in order to account for the different indicators, hence the two nodes on the right of Figure 2 called prevalence (1) and availability (2).

The use of SEM procedures has a number of advantages. Not only does it allow the identification of multiple constructs, it also allows models to be specified in terms of causal paths, as represented by estimating regression coefficients, between the constructs. Further, it considerably extends the ability to specify error structures and correlations between not only error terms, but also error variables and constructs themselves.

The structural component of the analysis involves path analysis, which may be defined as the process of estimating the coefficients of a set of linear structural equations (i.e. regressions) representing the predictive relationships hypothesised by the investigator.

Rather than estimating each equation separately, the SEM approach considers the model as a system of equations and estimates all the structural coefficients simultaneously. The ability of the model to estimate parameters simultaneously represents a considerable advance over earlier types of path analysis in which it was only possible to estimate parameters sequentially.

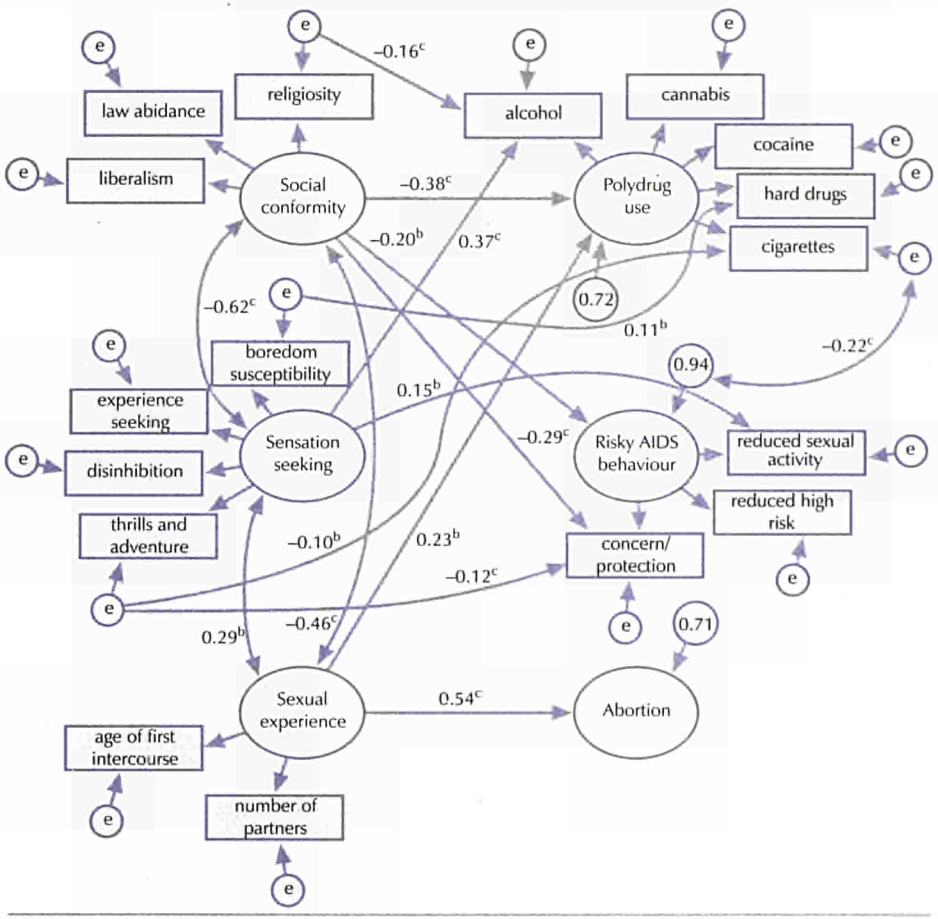
This greater modelling potential represents the use of data other than official statistics to derive drug-prevalence information, and to model behaviour in relation to drugs generally. The shift is driven by the recognition that prevalence *per se* is not precisely what is at the heart of drug addiction. Studies of personal behaviour allow a better description of many important facets of addiction prevalence.

An example of EQS

In a EQS structural-equation model every variable, be it latent or manifest, is defined as either independent or dependent. Dependent variables are those that can be expressed as a structural regression function of other variables. Independent variables are never regressed on each other. The two types of variables are then related by a set of structural equations. The analysis is usually conducted using the EQS structural-equation programme (Bentler, 1986).

The Stein et al. (1994) study illustrates the use of structural-equation modelling as a complex multi-step procedure. A stable, well-fitting and plausible factor structure was established from preliminary confirmatory-factor analyses. Here the 13 measured variables used in the initial analysis reflected the eight hypothesised latent variables in a reliable manner (see Figure 3). Further analyses were then undertaken using chi-square difference tests on various more restricted models. Specifically, the concordance, or similarity, among the substantive latent variables, once specific drugs have been

Figure 3: Significant regression paths among latent and manifest variables in the structural-equation model



Notes: [a]Regression coefficients are standardised, and residual variances are in circles
[b]p <0.01
[c]p <0.001
Source: Stein et al. (1994)

partialled out, are tested by setting restrictions on the general model by constraining pairs of correlations between latent variables at 1.00. These analyses are performed separately for each pair of constructs to test whether they are assessing identical traits. In other words, even though it is hypothesised on conceptual grounds that four substantive factors will be found, it is possible that fewer factors may account for the data. For instance, frequency and quantity factors may be measuring an identical tendency to use drugs.

The goodness-of-fit of the models is assessed in two ways: through use of a chi-square test of the overall fit of the model; and through use of the Bentler-Bonett normed fit index (NFI) statistic, which takes sample size into account. Acceptable values of the NFI should equal at least 0.90 (Bentler and Bonett, 1980).

Second-order factors are also tested. A second-order factor is a latent variable that is assumed to reflect an underlying relation or commonality among the primary latent variables. In particular, it is of interest to determine if the frequency and quantity latent variables may themselves result from a general drug-use second-order factor. Furthermore, another second-order general factor of drug-use and abuse is hypothesised and tested to account for all four of the substantive latent variables. If such a general, second-order factor were supported by the data, this would indicate that all types of drug-using behaviours, as well as negative consequences of drug-use, are generated by a pervasive disposition for drug involvement spanning both use and abuse.

After completion of the preliminary assessments of the factor structure, regression effects among the latent variables are examined. In particular, drug-use frequency and drug-use quantity are used to predict disruptive drug use and subjective problem drug use in order to assess the relative contribution of each predictor variable to these negative or adverse outcomes. These relations represent one of the main focal points of interest in the study because they represent associations among general drug factors after the influence of using specific drugs has been controlled. Finally, a model is tested that includes the negative drug consequences latent variable as part of the confirmatory-factor analysis. Then the negative drug consequences latent variable are included as a consequent variable in the regression model.

Specific problems arising from drug use are reflected in actual adverse consequences reported by the subjects. These variables include car accidents while under the influence of alcohol or other drugs, arrests or convictions for driving while intoxicated from alcohol or other drugs, and drug-law violations for selling or possessing drugs. The prediction of these specific drug-related behaviours is then based on the substantive latent variables identified in previous analyses. These problem variables are not included in the original multi-trait multi-method assessment because they do not relate specifically to separate drugs.

Both normal-theory maximum-likelihood (ML) and arbitrary-distribution theory (asymptotically distribution free; Browne, 1984) generalised least-squares (AGLS) estimation methods were used to estimate preliminary solutions for the model. Because many of the variables were non-normally distributed (normalised multivariate kurtosis estimate for the 13 variables = 319.75), the linearised AGLS (Bentler, 1983; Bentler and Dijkstra, 1985) solution was used in all subsequent analyses.

Results supported the construct validity of the hypothesised latent factors. Although highly correlated, drug-use frequency and drug-use quantity were not interchangeable: quantity was a more powerful predictor of disruptive and problem drug use. A second-order factor of general drug use and abuse accounted for these first-order constructs.

LISREL example

In contrast to an EQS model, a linear structural-relations (LISREL) model defines a set of latent response variables. These latent response variables depend on a set of latent explanatory variables. The relationship between the latent response and explanatory variables is defined in what is known as the *structural model*. The latent explanatory variables cannot be observed directly. However, some manifest response variables and manifest explanatory variables are observed. These are linked to the latent variables, both response and explanatory, through what is known as the *measurement model*. The specification of both the structural and measurement models is left to the user, and care must be taken to define an overall model which is both plausible and calculable. This approach formally develops a model in which the measurement model is separate from the structural relations between the constructs or latent variables.

Specifically using the idea that latent variables could be used generally for all constructs required by the models, Frischer et al. (1991) used two indicator, or manifest, variables for each of the latent variables of interest. Frischer et al. (1991) used data obtained as part of a cross-national study sponsored by the World Health Organisation, and a representative sample of current Glasgow IDUs were recruited from 'in-treatment' and 'out-of-treatment' sites. For the purposes of this study, in-treatment sites were defined as agencies whose principal aim is to modify drug-taking behaviour. A total of 536 IDUs were eligible for participation in the study. The LISREL programme was used to construct measurement and structural models relating to injecting drug use (Joreskog and Sorbom, 1989).

Subjects were included in the Frischer et al. (1991) analysis who had injected drugs in the two months prior to interview and had not already been interviewed for the study in the current calendar year. In addition, those recruited from in-treatment sites were only eligible if their current episode of treatment had begun within the previous four weeks. Each participant joining the study was told the name of the interviewer and the purpose of the study – to improve understanding of the drug-taking and sexual behaviour of IDUs in relation to their HIV seroprevalence. They were informed of the confidentiality of information to be obtained and the voluntary nature of the study. One hundred and seventy IDUs were recruited to the 'in-treatment' subset from 13 different agencies which together represent most of the drug-treatment capacity in Glasgow. The remaining 333 IDUs were recruited from 12 'out-of-treatment' settings, selected both to provide geographic spread and to maximise the likelihood that a proportion of the sub-set would include drug injectors who had no contact with drug-user treatment or harm-reduction agencies. The majority of respondents (84 %, N = 423) had injected in all of the six months prior to interview. The mean age of first drug use (by injection) was 17.3 years, and the mean length of drug-use injection was 6.8 years. The three drugs most commonly injected by respondents were buprenorphine, heroin and temazepam. Sample characteristics used as control variables in the analysis are described in the section on the measurement model below and shown in Table 1.

For the measurement model, ten features of injecting drug use were identified as being the most frequent topics of research (see Table 1) from a literature search which focused on injecting drug use and HIV-related risk practices. The measurement component of a SEM model consists of a number of underlying or latent variables that account for the

Table 1: Variables in a LISREL measurement model

Latent variable	Indicators	Correlation
Injecting drug use	A. Number of drugs injected B. Frequency of injection of all drugs	0.728
Sharing of injecting equipment	A. Frequency of equipment given by other B. Frequency of equipment given to others	0.443
Sexual activity (not prostitution)	A. Number of sexual partners B. Frequency of sex	0.318
Prostitution	A. Number of clients B. Frequency of all sex	0.711
(Current) awareness of AIDS	A. Frequency of talking about AIDS B. Number of routes of infection mentioned	0.270
Harm reduction (sexual activity and drug use)	A. Number of changes made B. Number of changes maintained	0.777
Travel (12 years prior to interview)	A. Number of places injected outside Glasgow B. Number of places had sex outside Glasgow	0.381
Prison experience	A. Frequency of injecting/sharing in prison B. How many nights spent in prison	0.275
Income	A. Scale of main source of legal income B. Percentage of income from illegal sources	0.295
Treatment (lifetime)	A. Number of types of treatment B. Frequency of treatment	0.852

Note: except where noted, indicators refer to behaviour in the six months prior to interview

common relationship among a number of observed variables. Previous work has indicated that it is inappropriate to interpret a latent construct as a measure of substantive importance in relation to its indicators where all its squared multiple correlations are less than 0.55 (Cuttance, 1987). These parameters are a measure of the validity of the measurement model.

For the structural model, all paths from all to all constructs were included in this particular model, an attempt was made not to pre-judge any issue of cause or effect, and the model was subsequently 'trimmed' of non-significant links (bi-directionally). As in all regression analyses, the effect of any independent variable on any dependent variable is tested independently of the variance shared with other variables. All parameters were estimated using the method of maximum likelihood.

Additionally, seven exogenous variables were included in the model. Each of these variables was measured by a single indicator and was constrained in the model to be predictive of, but not predicted by, the latent variables. These variables were socio-demographic in nature. Again following a global strategy, all relationships between the constructs and the exogeneous variables were controlled when estimating the path coefficients between the set of constructs.

Conclusions

The structural-equation-model procedures when brought to bear on very complex behaviours such as drug-addiction prevalence, intensity and problems can afford a richness of models to supply answers that are over-powering. The multiplicity of causal link paths between constructs that can only be loosely related to their manifest indicator measures, coupled with the ability to allow for measurement models that have correlated error variables, can generate models that are far from parsimonious (see e.g. Figure 2 from Stein et al., 1994). Such models tackle a particularly complex set of inter-relationships. It is not clear that such extensive modelling actually clarifies the relationships between the observed phenomena.

Frischer et al. (1991) report on a study that carries the idea of estimating structural paths to a logical conclusion, implying that any regression-analysis model could be in fact replaced by a path analysis between latent variables within a structural-equation model. A study that assesses relationships using several regression models (as did the original path models) could take advantage of the larger structural-equation model by incorporating all such analyses, if using a common or overlapping set of variables, into a single model.

Stein et al. (1988) report a study that uses structural-equation modelling extensively, with the frequency/quantity issue as central. The study was designed to ascertain whether measures of frequency and quantity are essentially interchangeable as predictors of problem or disruptive use of drugs in an older sample. They further examined the interrelations among different ways to assess drug use, self-reported problems with use, and specific negative outcomes resulting from drug use.

An explicit measurement model was developed by Person et al. (1977) in which it was hypothesised that each indicator, regardless of the method of measurement, was a monotone function of heroin abuse and an error term unique to the indicator. Tests of the measurement model, which derive from non-metric confirmatory-factor analysis, using non-metric statistics (McDonald, 1967) revealed an adequate fit of the model to data collected at four different time periods.

Subsequently, Woodward (1978) applied linear structural-equation models to test statistically a variety of measurement models of heroin-abuse indicators drawn from medical treatment and law-enforcement domains, addressing a variety of reliability/validity issues. Within a given quarter of the year, two alternative measurement models were contrasted. The first model allowed for factors defined by variables having a common method of measurement, and the second model postulated a single heroin-abuse construct underlying all indicators. The single-construct model could be accepted as plausible while the method-factor model could not.

While these early studies of the reliability and validity of indicator-based estimates of heroin abuse are important, they can be considered as only an initial step in the right direction. Because indirect methods such as factor analysis involve so many untenable assumptions, inaccuracies and influences beyond the control of the researcher, their practical utility remains constantly in doubt. In theory, the factor-analytic method works

well (for a simulation, see Person et al., 1976), but application to SMSA aggregate data remains a problem because of potentially shifting reporting bases underlying the indicators, distributional problems of the data, and other practical violations of assumptions. The same, however, can be concluded of the other major indirect estimation procedure, the capture–recapture method.

In conclusion, SEM models have been used successfully and have great potential to be used again in helping researchers to understand the complex relationships between various factors and in understanding how underlying theoretical concepts can be fitted into models of drug prevalence. As such, these methods are excellent tools for model generation and for stimulating discussion. That said, care should be taken not to take the results of such models too literally: these methods should always be viewed as powerful explorative tools.

References

Bentler, P. M. (1990). 'Comparative fit indexes in structural models'. *Psychological Bulletin* 107: 238–46.

Bentler, P. M. (1986). *Theory and implementation of EQS: a structural equations program*. Los Angeles, CA: BMDP Statistical Software.

Bentler, P. M. (1983). 'Some contributions to efficient statistics in structural models: specification and estimation of moment structures'. *Psychometrika* 48: 493–517.

Bentler, P. M., Bonett, D. G. (1980). 'Significance tests and goodness of fit in the analysis of covariance structures'. *Psychological Bulletin* 88: 588–606.

Bentler, P. M., Dijkstra, T. (1985). 'Efficient estimation via linearization in structural models', in Krishnaiah, P. R. (ed.) *Multivariate analysis*- VI. Amsterdam: Elsevier Science.

Browne, M. W. (1984). 'Asymptotically distribution-free methods for the analysis of covariance structures'. *British Journal of Mathematical and Statistical Psychology* 37: 62–83.

Campbell, D. I., Fiske, D. W. (1959). 'Convergent and discriminant validation by a multitrait–multimethod matrix'. *Psychological Bulletin* 56: 81–105.

Clayton, R. R., Ritter, C. (1985). 'The epidemiology of alcohol and drug abuse among adolescents'. *Advances in Alcohol and Substance Abuse* 4: 69–97.

Cuttance, P. (1987). 'Issues and problems in the application of structural equation models', in Cuttance, P., Ecob, R. (eds) *Structural modelling by example: applications of educational, sociological and behavioural research*. Cambridge: Cambridge University Press.

Demaree, R. G., Simpson, D. D., Joe, G. W. (1978). 'Recidivism and indices of drug abuse treatment effectiveness'. *Drug Alcohol Dependency* 3: 221–2.

Fitzgerald, J. L., Mulford, H. A. (1984). 'Factors related to problem-drinking rates'. *Journal of Studies in Alcohol* 45: 424–32.

French, J. F. (1977). 'Prevalence of heroin use in New Jersey'. Unpublished report of the Department of Health, Alcohol, Narcotic and Drug Abuse Unit, State of New Jersey.

Frisher, M., Bloor, M., Finlay, A., Goldberg, D., Green, S., Haw, S., McKeganey, N., Platt, S. (1991). 'A new method of estimating prevalence of injecting drug use in an urban population: results from a Scottish city'. *International Journal of Epidemiology* 20: 997–1000.

Greenwood, J. A., Crider, R. (1978). *Estimated number of heroin addicts in 1977*. Rockville, MD: National Institute on Drug Abuse.

Harvey, R. J., Billings, R. S., Nilan, K. (1985). 'Confirmatory factor analysis of the Job Diagnostic Survey: good news and bad news'. *Journal of Applied Psychology* 70: 461–8.

Huba, G. J., Newcomb, M. D., Bentler, P. M., Pandina, R. J. (1982). 'Comparison of adolescent drug use causal models controlling for response factors'. Paper presented at the meeting of the Western Psychological Association, Sacramento, CA, April.

Joreskog, K. G. (1971). 'Statistical analysis of congeneric tests'. *Pyschometrika* 36: 109–33.

Joreskog, K. G., Sorbom, D. (1989). *LISREL 7 users' reference guide*. Mahwah, NJ: Lawrence Erlbaum Associates Inc. in association with Scientific Software Inc.

Makela, L., Simpura, J. (1985). 'Experiences related to drinking as a function of annual alcohol intake by sex and age'. *Drug and Alcohol Dependence* 15: 389–404.

McDonald, R. (1967). *Nonlinear factor analysis*. Psychometrika Monograph 15. Chicago, IL: University of Chicago Press.

Minichiello, L. (1974). *Indicators of intravenous drug use in the United States 1966–1973*. Rockville, MD: National Institute on Drug Abuse.

Newcomb, M. D., Bentler, P. M. (1986a). 'Cocaine use among adolescents: longitudinal associations with social context, psychopathology, and use of other substances'. *Addictive Behaviors* 11: 263–73.

Newcomb, M. D., Bentler, P. M. (1986b). 'Cocaine use among young adults'. *Advances in Alcohol and Substance Abuse* 6: 73–96.

Person, P., Retka, R. L., Woodward, J. A. (1977). *A method of estimating heroin use prevalence*. Rockville, MD: National Institute on Drug Abuse.

Person, P., Retka, R. L., Woodward, J. A. (1976). *Toward a heroin problem index – an analytic model for drug abuse indicators*. Rockville, MD: National Institute on Drug Abuse.

Riley, W. J., Wagenfeld, M. O., Sonnad, S. (1981). 'Triangulated investigating: an approach to the estimating of the extent of heroin use'. *Addiction* 16: 97–108.

Sadava, S. W. (1985). 'Problem behaviour theory and consumption and consequences of alcohol use'. *Journal of Studies on Alcohol* 46: 392–7.

Sadava, S. W., Secord, M. J. (1984). 'Problem drinking in context'. Paper presented at the International Congress of Psychology, Acapulco, Mexico.

Schlenger, W. E., Greenberg, S. W. (1978). *Analysis of intermediate correlates of heroin usage*. Rockville, MD: National Institute on Drug Abuse.

Schmitt, N., Stults, D. M. (1986). 'Methodology review: analysis of multi-trait multi-method matrices', *Applied Psychological Measurement* 10: 1–22.

Stacy, A. W., Widaman, I. F., Hays, R., DiMatteo, M. R. (1985). 'Validity of self-reports of alcohol and other drug use: a multi-trait multi-method assessment'. *Journal of Personality and Social Psychology* 49: 219–32.

Stein, J. A., Newcomb, M. D., Bentler, P. M. (1994). Psychosocial correlates and predictors of AIDS risk behaviour, abortion, and drug use among a community sample of young adult women'. *Health Psychology* 13: 308–18.

Stein, J. A., Newcomb, M. D., Bentler, P. M. (1988). 'The relationship of gender, social conformity, and substance use: a longitudinal study'. *Bulletin of the Society of Psychologists in Addictive Behaviours* 5: 125–38.

Stein, J. A., Newcomb, M. D., Bentler, P. M. (1987). 'An 8-year study of multiple influences on drug use and drug use consequences'. *Journal of Personality and Social Psychology* 53: 1094–105.

Stoloff, P., Levine, D. B., Spruill, N. (1975). *Public drug treatment and addict crime*. San Francisco, CA: Public Research Institute.

Woodward, J. A. (1978). 'Structural equation models of drug abuse indicators'. Paper presented at the National Drug Abuse Conference, Seattle, WA, April.

Woodward, J. A., Retka, R. L., Ng, L. (1984). 'Construct validity of heroin abuse estimators'. *International Journal of the Addictions* 19: 93–117.

Modelling the health consequences of drug use

Part IV

INTRODUCTION

This part concentrates on modelling adverse health consequences of problem drug use, possibly the main reason for policy and intervention in the field of drug use. One of the most important health consequences of illicit drug use is infectious diseases such as AIDS. At the same time the field of infectious-diseases epidemiology is one of the public-health areas that may have benefited most from modelling studies and may therefore serve as an example for approaches that could be applied to the drugs field.

In Chapter 10, Kretzschmar follows a different approach from the chapters in Part III in starting from an explicit problem to be solved rather than from the methods to be used. She discusses the main disease-specific parameters that have been much used in models of the sexual transmission of HIV, but less so in models of injecting drug use or other drug-related infections. Kretzschmar defines dynamic models and introduces the concepts of a core group and a mixing matrix before considering the influence of partnership duration in pair-formation models and network models, where each individual and each contact between two individuals are modelled explicitly. The author argues that these models can easily be adjusted for hepatitis B and C. Techniques discussed in the chapter could be used to answer questions such as how well the future course of these drug-related disease epidemics can be predicted, what the effects of behaviour change would be on these diseases, and what prevention strategies are to be recommended.

Jager and his colleagues give a structured overview of developing multinational scenario analyses in Chapter 11. The authors argue that the drug problem is well suited to this form of analysis as the issue is serious and important, complex, multinational, multi-disciplinary and in need of information to inform policy decisions. The main elements of a scenario study are the development of a conceptual model, the translation into mathematical models, the performance of a basic analysis and the construction of scenarios. A preliminary conceptual model for the drugs field is presented, besides examples of conceptual models for AIDS and blood supply that have already been applied at EU level.

CHAPTER 10

MODELLING INFECTIOUS DISEASES AND OTHER HEALTH CONSEQUENCES: PREDICTING FUTURE MORBIDITY CONSEQUENCES AND SPREAD OF HIV, HEPATITIS B AND HEPATITIS C

Mirjam Kretzschmar

A very high prevalence of HIV, hepatitis B virus (HBV) and hepatitis C virus (HCV) is found in populations of injecting drug users (IDUs) due to a combination of high-risk behaviour and high transmission rates via sharing of contaminated injection equipment. In many European countries, HIV incidence among IDUs is alarmingly high and prevalence is rising explosively. In southern European countries, the largest fraction of new HIV infections is among IDUs. While the incidence among homosexual men has stabilised and prevalence seems no longer to be rising, among IDUs prevalence is still increasing. For HBV and HCV prevalence is even higher. Although hepatitis may not be as serious an illness as AIDS, it can nevertheless lead to major health problems and in a small fraction of cases can lead to liver cancer. A sustained prevalence of carriers of the hepatitis virus also constitutes a source of infection for the entire population. These infectious diseases are therefore important consequences of drug misuse.

Mathematical models are used to attempt to understand the quantitative relationships between the characteristics of a specific disease, the contact patterns in the population within which it is transmitted, and the resulting incidence and prevalence of the disease. These insights can then be used to support the interpretation of epidemiological data, help decide which data are needed to be able to assess an epidemiological situation, and help design effective prevention measures. To extend disease modelling in this way requires further steps and information. This could, for example, include costs of prevention programmes, or the quality of life for given levels of prevalence and incidence over a certain time period. It is clear that the time scale in which the spread of the epidemic takes place, and the time scale in which prevention programmes are designed, play an important role in the outcome of the evaluation of prevention programmes.

The disease-specific parameters of importance for modelling are, among others, the probability of transmission per contact, where a contact is defined as an event in which transmission can possibly take place; the duration of the latent period and the infectious period; the fraction of infections that are asymptomatic, but are infectious to others; the recovery rate. Which of these parameters are incorporated into the model depends

on the characteristics of the disease being modelled and on the level of detail that is needed to properly describe transmission and the required outcome variables. For example, in modelling the spread of HIV it might be useful to take a number of stages of infection into account distinguished by their CD4+ counts, even if these stages do not play distinct roles in the transmission process. Concerning the contact patterns that underlie transmission, these are obviously determined by the transmission route of the disease in question. For a sexually transmitted disease it matters whether spread is to be modelled by homosexual or heterosexual contact; furthermore, the number of sexual partners per time unit (e.g. year), the duration of the sexual partnerships, the overlap in time of partnerships or 'concurrency', and the heterogeneity with respect to these variables in the population are determinants of the contact patterns. Similarly, for IDUs the contact structure is constituted by the social contacts of drug users within which injecting equipment is shared, such as anonymous contacts in shooting galleries, or longer-lasting relationships with friends and sexual partners. In this case, the existence of two different transmission routes for the infection, namely the needle-sharing and the sexual routes, makes the contact patterns even more complex.

The aim of this chapter is to give an overview of the work that has been done in the area of infectious-disease modelling, and more specifically in modelling the spread of HIV via sexual contact and needle sharing. The relevant questions and approaches are identified and the implications for future work to investigate the spread of HIV, HBV and HCV in populations of IDUs are summarised. Finally, the findings are evaluated in the light of questions posed at EU level in the context of this monograph. The focus is on so-called dynamic models, that is, models that explicitly describe the transmission process and the resulting dynamics of the spread of an infection, as opposed to statistical models, such as the back-calculation method, that are used for data-based projections. For an application of the latter methodology to data from the European Union, see Downs et al. (1997) and for a fuller description of this method, see Chapter 6.

Models for HIV, HBV and HCV

Since the beginning of the AIDS pandemic in the early 1980s, an enormous effort has been made in devising mathematical models to predict the further spread of HIV. Rather than attempting to give a complete overview of the existing models, this chapter focuses on the relevant modelling questions that have been discussed and that are essential not only for HIV, but for any infectious disease that spreads via sexual contact or via shared injecting equipment. It should be said that the large majority of all models investigated so far have focused on the former transmission route, and many of these have dealt specifically with the transmission group of homosexual men. Only very few models have addressed questions pertaining to transmission via needle sharing. These models will be discussed in the following section. For the spread of HBV and HCV, to the author's knowledge only very few models exist and will be mentioned briefly at the end of this section.

Even before AIDS had become the dominant problem in the area of sexually transmitted diseases (STDs) it was recognised that the heterogeneity in sexual behaviour of a

population plays an important role in determining the patterns of spread of an STD. In modelling prevention strategies for gonorrhea, Hethcote and Yorke (1984) introduced the concept of a 'core group', a small sub-group of the population engaging in very high levels of sexual activity, in other words, a high rate of partner change. With the advent of AIDS, attention to the incorporation of a detailed description of sexual behaviour increased greatly, and a number of modelling approaches developed. The basic idea was to distinguish different activity levels in the population and to define a 'mixing matrix' describing the contact rates within and among sub-groups of different sexual activity (see e.g. Anderson and May, 1991; Blythe and Castillo-Chavez, 1989; Jacquez et al., 1989). It was then possible to investigate how mixing between sub-groups determines the epidemic spread, and how prevention focused on specific sub-groups affects incidence and prevalence. However, it turned out to be very difficult to collect reliable data with which to estimate the parameters of the mixing matrix.

A different approach in modelling focused on the influence of partnership duration on the epidemic spread. In the above-mentioned mixing models, it is implicitly assumed that every contact is with a new individual and there are no longer-lasting partnerships. In contrast, in the so-called 'pair-formation models', first introduced in the context of STDs by Dietz and Hadeler (1988), the formation and separation of possibly long-lasting partnerships between pairs of individuals is explicitly taken into account. Especially for the spread of STDs in heterosexual populations, partnership duration cannot be neglected as a determining parameter. Although conceptually pair-formation models are very appealing, they have the disadvantage that including more structure into the model, such as sub-groups with different levels of sexual activity or the possibility to have more than one partnership at a time, rapidly increases the complexity of the model.

A further step towards a detailed description of sexual behaviour is made by 'network models' (Kretzschmar et al., 1994). These are models that try to capture the essential features of sexual networks by describing the population explicitly as a collection of individuals and their relationships. In contrast to the types of models discussed above, which are implemented as systems of differential equations, network models are inherently stochastic and require simulation by Monte Carlo methods for obtaining results. In other words, except in very simple cases, it is not possible to derive analytic 'parameter-value-independent' results, but it is necessary to run and analyse large numbers of simulations and to apply statistical techniques as if analysing real data. The advantage of this procedure is that the simulations are easily comparable with epidemiological data and data obtained by sexual-behaviour surveys. It has been shown (Morris and Kretzschmar, 1997) that a description of sexual partnerships in terms of numbers per year and duration is not sufficient to predict the spread of an STD, but also that the possible overlap and therefore concurrency has a large effect on the speed and size of an epidemic. Simultaneous, longer-lasting partnerships cannot be modelled with mixing models and only in a very limited degree with pair-formation models; they arise naturally, however, in network models.

Although these modelling approaches were developed with the aim of describing the sexual-contact structure in a population, it is obvious that the same questions will play

a role when describing the transmission of a disease via sharing injection equipment. In that case an important determinant of the spread of an infectious disease is the heterogeneity in behaviour, that is, the possible existence of sub-groups of the population with high-risk behaviour and their mixing patterns with the rest of the population with lower-risk behaviour. It is important to consider whether the sharing of needles takes place within long-lasting relationships, that is, with friends or sexual partners, or whether it takes place more or less randomly as is the case in shooting galleries. If longer-lasting relationships are the basis for risk behaviour and disease transmission, the question will be how best to describe the networks these contacts constitute. In short, modelling techniques that have been developed to describe the spread of an infection via sexual contacts, can – with the proper adjustments – be used for modelling transmission via needle sharing. Of course there are additional factors to be considered that will be discussed in the following section.

To describe a specific disease in terms of model variables and parameters the relevant disease stages, their duration and the transmission probabilities related to those stages must be specified. With respect to HIV/AIDS the discussion has focused on its long incubation period and the varying infectivity during that period. Using statistical models, Longini et al. (1989a), Longini et al. (1996) and Hendriks et al. (1998) have estimated the time from infection to development of AIDS for homosexual men and IDUs, respectively. For the transmission of the infection and therefore its epidemiology, the probability of transmission per contact in the various stages of infection is important. As Longini et al. (1989b) and Jacquez et al. (1994) have argued convincingly, data and modelling results support the idea that infectivity is very high in the first acute phase of infection, very low during the long asymptomatic phase, and higher again when the first symptoms of AIDS become manifest. Blythe and Anderson (1988) have investigated how variable infectivity affects the spread of the epidemic in a mixing model; Kretzschmar and Dietz (1998) have investigated the effects of variable infectivity in models with long-lasting partnerships.

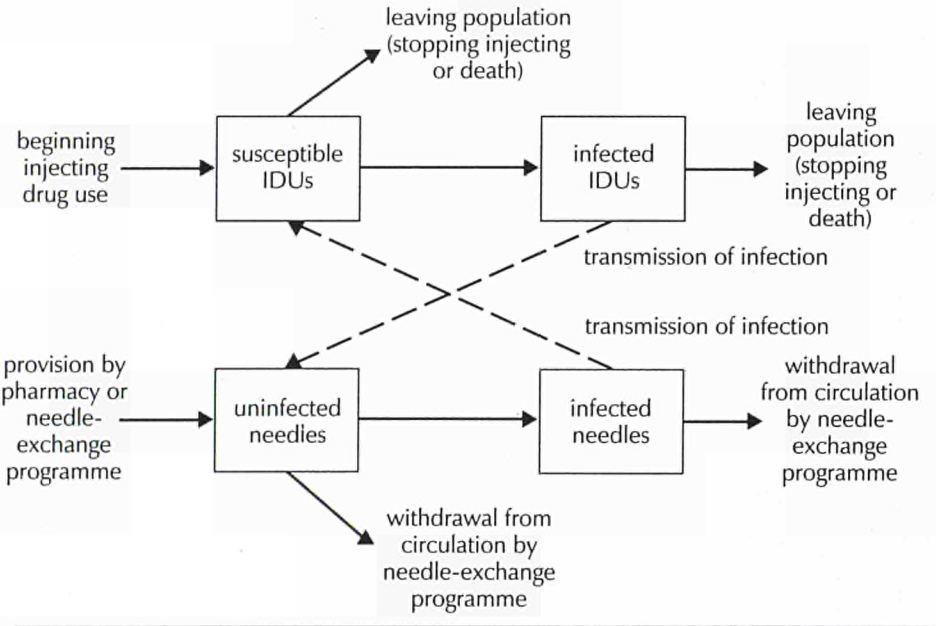
Although the prevalence of hepatitis B and C is extremely high in some risk groups, there has been surprisingly little work done on modelling the spread of those diseases. The only work to the author's knowledge is by McLean and Blumberg (1994) and by Williams et al. (1996a, 1996b). Both models are mixing models, as described above, and incorporate features specific to hepatitis B, such as the existence of an asymptomatic carrier stage. While the former focuses on modelling countries with high transmission rates, the latter investigates the cost effectiveness of immunisation programmes against hepatitis B in the UK. Furthermore, the latter model makes a distinction between transmission in the heterosexual and the homosexual populations, respectively, characterised by different values of the parameters describing sexual activity. The vaccination strategies discussed are antenatal screening, mass infant and adolescent vaccination programmes with different coverages, and vaccination targeted at genito urinary (GU) clinic attenders with various coverages.

Models for HIV in injecting drug users

Models that have been designed to describe the spread of HIV specifically in populations of IDUs are summarised in Table 1. The transmission via needle sharing adds a complication that is not present with sexual transmission, namely that transmission takes place via a 'vector', the infected needle. As a result, contacts can be directed in the sense that if a susceptible borrows a contaminated needle from an infected person, transmission can take place, while if an infected person borrows from a susceptible, transmission is impossible. Of course, there are also less clearly defined situations, for example when a group of IDUs successively uses the same injecting equipment, which is even more complex to incorporate explicitly into a model.

The first model for the spread of HIV in IDUs was introduced by Kaplan (1989). In this model the complications of the transmission route were taken into account by explicitly describing the fraction of infected needles that circulated in the IDU community (see Figure 1). The risk of an IDU becoming infected depended on the fraction of infected needles, which in turn could be influenced by increasing the inflow of clean needles via a needle-exchange programme. It was possible to evaluate the effects of such a programme in a very short time and to show that needle exchange significantly reduced the risk of new infections. The Kaplan model has been extended to include a more complex description of transmission probabilities (Greenhalgh and Hay, 1996) and heterogeneity in the risk behaviour of IDUs (Greenhalgh, 1996, 1997).

Figure 1: Kaplan model



Source: Kaplan (1989)

Table 1: Summary of models and their data needs and possible application

Model	Output	Applicability	Data demands	Why not applied more?	Added value	Geographic area
Kaplan (1989)	estimate of effectiveness of needle-exchange programmes in reducing HIV incidence	applicable to cities with homogeneous mixing (shooting galleries or similar settings)	HIV prevalence, needle-exchange rates, fraction of infected needles circulating	only useful in combination with needle testing	objective tool for evaluating needle-exchange programmes, easy implementation	New Haven, CT
Greenhalgh and Hay (1996)	as above	homogeneous mixing, but different risk levels	differences in risk behaviour, probabilities of needles being infected	only theoretical results	refinement of Kaplan model for heterogeneous populations	not specified
Blower et al. (1991)	HIV prevalence for different transmission groups	cities with large IDU populations	needle-sharing behaviour, sexual-behaviour data	many parameters to be estimated	implementation not too difficult, includes sexual transmission	New York City, NY
Capasso et al. (1995)	threshold theorem (R_0)	?	?	only theoretical results	?	not specified
Gani and Yakowitz (1993)	expected number HIV infected as function of time	small groups of IDUs	detailed data on needle-sharing behaviour	too complex, too many assumptions	submodel for local spread of infection	not specified
Peterson et al. (1990)	HIV prevalence	cities	detailed data on needle-sharing behaviour, HIV progression	very complex, detailed data not available	tool for evaluation of behaviour changes and targeted prevention programmes	not specified
Kretzschmar and Wiessing (1998)	yearly HIV incidence under various prevention scenarios, HIV prevalence	small populations (cities) with small-scale mixing (sharing with friends/partners)	detailed behavioural data (rate of needle sharing, fraction of sharing from friends or strangers, etc.)	very complex, detailed data not available	tool for evaluation of behaviour changes and targeted prevention programmes	Arnhem, Rotterdam, the Netherlands
Iannelli et al. (1997)	projections of HIV and AIDS incidences	countries, split up by transmission groups	AIDS incidence among IDUs	similar models applied for HIV in other transmission groups	mathematical theory about this type of model available	Italy, especially Latium region

Note: ? indicates unknown

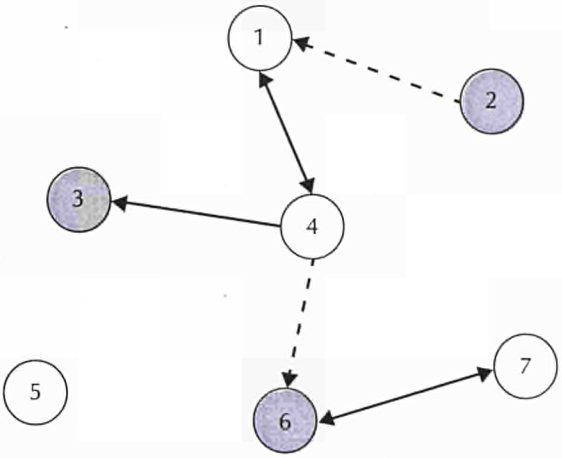
Blower et al. (1991) designed a model to describe the situation of IDUs in New York City. The model incorporated transmission via needle sharing, via heterosexual contact and via vertical transmission from mother to unborn child. Furthermore, differences in risk levels were taken into account by distinguishing between IDUs who share needles with strangers and others who share with 'buddies'. The model consisted of a system of 34 ordinary differential equations with a large number of parameters. One of the results obtained was that the stabilisation of prevalence that had been observed in New York City could be explained as the natural course of the epidemic without any behaviour changes. Also, the stabilisation of prevalence in the model was only temporary and was followed by a renewed increase in prevalence that was driven by heterosexual transmission. An important conclusion is that the stabilisation of prevalence may not be interpreted as being the consequence of behaviour changes without having more information about population structure and mixing parameters (Blower, 1991). This information, however, is very difficult to collect and the uncertainty in these parameters leads to extremely wide confidence intervals in the predictions of future incidence and prevalence.

The existence of groups of IDUs which exhibit risk behaviour is included in a deterministic model by Capasso et al. (1995) and in a stochastic model by Gani and Yakowitz (1993) and Yakowitz (1994). While the description of the group-forming process in a deterministic setting is questionable, in the stochastic setting it quickly leads to a very complex model. Thus as yet there appear to be few epidemiologically relevant results and future research remains to be done.

Peterson et al. (1990) and Kretzschmar and Wiessing (1998) investigated stochastic, individual-based models in which the contact patterns of IDUs can be described in much detail. The model of Peterson et al. consists of three sub-models: one is a model of HIV progression in an infected individual that distinguishes five stages of infection; the second describes the heterogeneity of injecting drug use within the population, distinguishing monthly, weekly and daily injectors and drug users in jail or in treatment programmes; the third sub-model describes the contact patterns in the IDU community and distinguishes between the acquaintance network of an IDU, individual strangers and shooting galleries. Transition probabilities are then defined for the various states an individual can be in and Monte Carlo simulations are performed. The authors investigated the effect of the social-network structure on the epidemic and found a far slower increase in prevalence in populations where sharing took place in acquaintance networks as opposed to random mixing in shooting galleries. Again, the prediction for prevalence in all cases displayed a wide variability, here as a consequence of the stochastic nature of the processes. Finally, the authors showed that the effectiveness of intervention is highly dependent on the prevalence level of the population; in a low-prevalence setting prevention can potentially be much more effective because it begins before the start of the epidemic.

In Kretzschmar and Wiessing (1998) the effect of different prevention strategies on the incidence of HIV in IDU populations is compared. The model distinguishes between needle sharing within long-lasting relationships (friends/partners) and sharing with individual strangers (see Figure 2). Furthermore, it is assumed that the infectivity of an infected individual is very high during the first six weeks of infection and very low during the long asymptomatic

Figure 2: Network of IDUs



Notes: Individuals 1, 4, 5 and 7 are susceptible; individuals 2, 3 and 6 are infected.
——— = steady, long-term relationships (buddies)
– – – – = incidental borrowing from strangers
Arrows show the direction of passing on a needle. Transmission can only take place if the individual at the base of the arrow is infected and the individual at the point of the arrow is susceptible.
Individuals 1 and 4 have a long-term relationship in which they regularly share needles. They are both uninfected, so cannot infect each other. Individual 4 also regularly lends needles to individual 3, but does not borrow any from individual 3. Therefore individual 4 cannot get infected by individual 3. However, individual 1 borrows a needle in an incidental contact from individual 2, and could get infected in that way. Individual 5 does not borrow needles at all and is not at risk of getting infected. Individual 7 borrows regularly from his infected partner (individual 6) and has a large risk of eventually becoming infected.
Source: Kretzschmar and Wiessing (1998)

phase afterwards. Values for behavioural parameters were estimated from surveys in two cities in the Netherlands. Analysing the results obtained with Monte Carlo simulations the authors investigated the effects of various contact patterns on the prevalence and incidence of HIV. They found that there was a clear threshold phenomenon for the sharing frequency, that is, that below a certain average sharing frequency the epidemic never takes off while above the threshold it quickly reaches high levels of prevalence. Also, the variability in the course of an epidemic as a consequence of the stochastic nature of events and the transmission via a contact network is large, which means that predictions about the future course of an epidemic are uncertain. In the comparison of prevention strategies, the authors concluded that focusing prevention on new injectors can possibly have a large effect on HIV incidence, while behaviour change of IDUs after a positive HIV test has no significant effect at a testing rate of 50 % per year. Finally, reducing the sharing frequency with strangers has a larger impact on the epidemic than the same reduction in the sharing within the stable buddy network.

Iannelli et al. (1997) examined several hypotheses about the dynamics of the HIV epidemic among injecting drug users in Italy by fitting a model to the existing data and using a least-squares approach to determine the best fit. The model is a variant of the classical susceptible, infected, removed (SIR) model, but includes variable infectivity depending on the time since infection. The authors conclude that the fit of the model to AIDS-incidence data can be improved with respect to a simple reference model by either assuming a reduction in risk behaviour over time, or assuming that there is a small sub-group in the population with high-risk behaviour. Furthermore, the assumption of constant infectivity turned out to be untenable, while variable infectivity seemed to be in agreement with the data, thus confirming earlier empirical findings and theoretical results.

Although none of the models discussed in this section is designed to describe the spread of hepatitis B and C, most models can be adjusted fairly easily to account for the specific properties of those diseases.

Future role of modelling

Many problems remain to be solved in appropriately modelling the social structure underlying the transmission of infectious diseases via needle sharing, and collecting data that will provide more quantitative information about contact patterns and risk behaviour is essential. In addition, little work has been done on analysing the interaction between needle-sharing networks and sexual networks. This section examines what is needed to make modelling successful, and discusses some research topics in which modelling can play a key role in investigating hypotheses and interpreting data.

A prerequisite in making any modelling effort successful will be the understanding of the contact patterns of IDUs, of their essential features and how these are related to the spread of an infectious disease. 'Contact pattern' here means a quantitative description of behaviour that relates to possible transmission events. For the sharing of injection equipment the following needs to be known: techniques of sharing such as borrowing used syringes, backloading; with whom this behaviour takes place, whether steady groups of IDUs, changing groups, steady sexual partners or strangers; how often risk behaviour takes place, that is, borrowing rates per month; the typical 'life history' of an IDU, in other words, how risk behaviour changes during an IDU's injecting career, how long do IDUs on average stay in the population of drug users. An important aspect of all these variables will be their heterogeneity: a population of IDUs is probably not homogeneous in its behaviour, but will contain sub-groups of individuals with very different behaviour. For example, there is a small core group of IDUs with high-risk behaviour including: high rates of injecting, high rates of sharing used equipment possibly in settings that encourage sharing with strangers. Other IDUs, who restrict themselves to sharing with their steady sexual partner, contribute much less to the spread of the epidemic. Similar aspects are important for the sexual transmission route. While for sexual behaviour large surveys have been conducted, the information about behavioural aspects of sharing injection equipment is still scarce.

A task for the modeller is to identify those aspects of risk behaviour that can be quantified and that determine the epidemiology of the disease in question. The aim is therefore to find a (preferably small) number of aggregated variables that satisfactorily describe risk behaviour, that are possible to measure in the field and that are sufficient for explaining epidemiological phenomena or even predicting the qualitative and quantitative development of an epidemic, although the latter might only be a theoretical possibility. It immediately follows that this information can be used to design prevention programmes because knowledge of the risk factors that have the largest influence on, for example, the incidence of a disease indicates which prevention measures will have the largest effect on incidence.

Furthermore, modelling can be used to test hypotheses about the parameters describing the course of an infection. Comparable to the discussion about the variable infectivity in HIV-infected persons, there might be questions related to the natural course of a hepatitis infection which could be investigated in a model. Reliable epidemiological or microbiological data are then needed in order to compare the fit of model results under different hypotheses with the data and then decide on which hypothesis produces the best fit with observations.

A very interesting study will be the comparison of the way different infectious diseases spread in the same population. In a population of IDUs, in which needle sharing and sexual behaviour are the transmission routes, how on the basis of disease-specific parameter values an epidemic of HIV, HBV and HCV proceeds can be compared. What is the equilibrium prevalence and incidence for these diseases in the same population? What degree of co-morbidity can be expected? How do similarities and differences depend on the underlying contact parameters? If good information were available about the prevalence and/or incidence of these three diseases, could conclusions be drawn about the underlying levels of risk behaviour? It is expected that there will be differences in the effectiveness of a given prevention strategy as a consequence of differences in disease-specific parameters such as transmission rates and the resulting variations in prevalence.

It has been shown that HCV is transmitted mainly via needle sharing (or more generally blood-to-blood contact), while HBV spreads via sexual contact and via needle sharing (van Ameijden et al., 1993). A comparison of prevalence and incidence of these two viruses in a population of drug users can therefore give some information about the interaction of the two transmission networks. This in turn would be very valuable information for the design of prevention measures. A certain fraction of new infections could be attributed to the two transmission routes, possibly depending on the risk behaviour with respect to these transmission routes, and prevention could be focused on the dominating form of transmission. For example, how important are prostitutes who inject drugs as bridges between the IDU network and the core group with high sexual activity?

It is also possible to link the incidence of HBV and HCV to the prevalence of injecting drug use. If the fraction of infections due to transmission via needle sharing, the fraction of asymptomatically infected persons or chronic carriers, and the prevalence of HBV

and HCV in IDU populations were known, then the number of IDUs in a country could be estimated. An assumption here would have to be that there was a stable endemic situation and that there was sufficient knowledge of the risk behaviour among IDUs to estimate the equilibrium prevalence.

In the context of a World Health Organisation (WHO) programme promoting vaccination against HBV infection, all countries are advised to vaccinate all newborns against hepatitis B from the year 1997 on. As in many European countries, prevalence is very low and restricted mainly to some high-risk groups, the cost-effectiveness of such a vaccination programme is questionable (cf. Williams et al. 1996a, 1996b). However, vaccination focused on risk groups such as IDUs may be very effective in reducing incidence as well as the prevalence of chronic carriers. In view of the fact that a female IDU who is infected with HBV might pass the virus on via vertical transmission, screening of pregnant women within that risk group is especially important. However, it is known that IDUs are a group that is difficult to reach so the effectiveness of vaccination will be limited by its coverage. Modelling can help to quantify the dependence of cost-effectiveness measures on coverage and other parameters of a vaccination programme.

Conclusions

The use of specific models by an organisation such as the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) or in the context of a concerted action of the European Union depends very much on the types of questions being investigated and possibly answered. Types of questions could be: how well can the future course of the HIV epidemic be predicted in different Member States of the EU? What are the effects of behaviour change on the prevalence of HIV, HBV and HCV in different European countries? Which prevention strategies are recommended for a country given the dominant type of risk behaviour in that country? What conclusions can be drawn about the prevalence and incidence of injecting drug use from the prevalence of HIV, HBV and HCV in different European countries?

Predictions about the future course of the HIV epidemic among IDUs in Europe have been made via the back-calculation method (Downs et al., 1997). That method has some shortcomings, however, that makes its results unreliable. Predictions are made on the basis of AIDS incidence, and HIV incidence can be estimated using knowledge about the incubation time distribution. It is not possible to incorporate effects of behaviour change explicitly into the model. Also, as the effects of treatment on the incubation time are not known, the estimates will in the future become even more uncertain than they are now.

Dynamic models, which explicitly model the transmission process, can be used to evaluate the effects of behaviour change and other prevention measures. However, dynamic models require much more detailed data for describing the transmission process, mainly data about risk behaviour (Blower and Medley, 1992). A start at collecting such data at European level has been made by Papaevangelou and Richardson (1995) and Richardson et al. (1993). Using dynamic models it is often advisable to begin with a very simple structure even if many simplifying assumptions have to be made, because

the results will usually allow a clear interpretation. A very good example is the model by Kaplan and Heimer (1992), which consisted only of two differential equations and did not take any of the complications of the contact structure into account. Nevertheless, Kaplan demonstrated the positive effects of a needle-exchange programme on HIV incidence so clearly that the programme was continued and new needle-exchange programmes were established in different cities in the United States. A summary of the models discussed above, their data needs and possible applications is shown in Table 1, above.

If more detailed questions about the effects of behaviour changes are to be answered, more complex models are needed that allow the implementation of a relevant contact structure. In this respect, individual-based simulation models seem to be the best choice at present. However, if modelling is to be closely related to empirical findings it will be necessary to collect detailed behavioural data. It is very important that modellers and epidemiologists work together from the start to avoid large gaps between model design and data format.

In conclusion, regardless of the amount of work that has been done in modelling the spread of HIV/AIDS, modelling the spread of infectious diseases via needle sharing and the specific problems connected with the spread of hepatitis B and C is still in its infancy. If models are to be used in an EU context, they should be developed according to the questions they are supposed to answer and in close interaction with social scientists and epidemiologists.

References

van Ameijden, E. C. J., van den Hoek, J. A. R., van den Mientjes, G. H. C., Coutinho, R. A. (1993). 'A longitudinal study on the incidence and transmission patterns of HIV, HBV and HCV infection among drug users in Amsterdam'. *European Journal of Epidemiology* 9: 255–62.

Anderson, R. M., May, R. M. (1991). *Infectious diseases of humans: dynamics and control*. Oxford: Oxford University Press.

Blower, S. M. (1991). 'Behaviour change and stabilization of seroprevalence levels in communities of injecting drug users: correlation or causation?' *Journal of AIDS* 4: 920–4.

Blower, S. M., Medley, G. (1992). 'Epidemiology, HIV and drugs: mathematical models and data'. *British Journal of Addiction* 87: 371–9.

Blower, S. M., Hartel, D., Dowlatabadi, H., Anderson, R. M., May, R. M. (1991). 'Drugs, sex and HIV: a mathematical model for New York City'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 321: 171–87.

Blythe, S. P., Anderson, R. M. (1988). 'Variable infectiousness in HIV transmission models'. *IMA Journal of Mathematics Applied in Medicine and Biology* 5: 181–200.

Blythe, S. P., Castillo-Chavez, C. (1989). 'Like-with-like preference and sexual mixing models'. *Mathematical Biosciences* 96: 221–38.

Capasso, V., Sicurello, F., Villa, M. (1995). 'Mathematical models for HIV transmission in groups of injecting drug users via shared drug equipment'. *Journal of Biological Systems* 3: 747–58.

Dietz, K., Hadeler, K. P. (1988). 'Epidemiological models for sexually transmitted diseases'. *Journal of Mathematical Biology* 26: 1–25.

Downs, A. M., Heisterkamp, S. H., Brunet, J. B., Hamers, F. F. (1997). '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* 11: 649–62.

Gani, J., Yakowitz, S. (1993). 'Modelling the spread of HIV among intravenous drug users'. *IMA Journal of Mathematics Applied in Medicine and Biology* 10: 51–65.

Greenhalgh, D. (1997). 'Bounds on the endemic disease levels, risks and basic reproductive numbers in heterogeneous models for HIV/AIDS amongst injecting drug users'. *Journal of Biological Systems* 5 (2): 151–80.

Greenhalgh, D. (1996). 'Effects of heterogeneity on the spread of HIV/AIDS among intravenous drug users in shooting galleries'. *Mathematical Biosciences* 136 (2): 141–86.

Greenhalgh, D., Hay, G. (1996). 'Mathematical modelling of the spread of HIV/AIDS amongst injecting drug users'. *IMA Journal of Mathematics Applied in Medicine and Biology* 14: 1–28.

Hendriks, J. C. M., Satten, G. A., van Ameijden, E. J. C., van Druten, J. A. M., Coutinho, R. A., van Griensven, G. J. P. (1998). 'The incubation period to AIDS in injecting drug users estimated from prevalent cohort data, accounting for death prior to an AIDS diagnosis'. *AIDS* 12: 1537–44.

Hethcote, H. W., Yorke, J. A. (1984). *Gonorrhea transmission dynamics and control*. Lecture Notes in Biomathematics Vol. 56. Berlin: Springer Verlag

Iannelli, M., Milner, F. A., Pugliese, A., Gonzo, M. (1997). 'The HIV/AIDS epidemics among drug injectors: a study of contact structure through a mathematical model'. *Mathematical Biosciences* 139: 25–58.

Jacquez, J. A., Koopman, J. S., Simon, C. P., Longini, Jr., I. M. (1994). 'Role of the primary infection in epidemics of HIV infection in gay cohorts'. *Journal of AIDS* 7: 1169–84.

Jacquez, J. A., Simon, C. P., Koopman, J. (1989). 'Structured mixing: heterogeneous mixing by the definition of activity groups', in Castillo-Chavez, C. (ed.) *Mathematical and statistical approaches to AIDS epidemiology*. Lecture Notes in Biomathematics Vol. 83. Berlin: Springer Verlag, 301–15.

Kaplan, E. H. (1989). 'Needles that kill: modelling human immunodeficiency virus transmission via shared drug injection equipment in shooting galleries'. *Review of Infectious Diseases* 11: 289–98.

Kaplan, E. H., Heimer, R. (1992). 'A model-based estimate of HIV infectivity via needle sharing'. *Journal of Aids* 5: 1116–18.

Kretzschmar, M., Dietz, K. (1998). 'The effect of pair formation and variable infectivity on the spread of an infection without recovery'. *Mathematical Biosciences* 148: 83–113.

Kretzschmar, M., Wiessing, L. G. (1998). 'Modelling the spread of HIV in social networks of injecting drug users'. *AIDS* 12: 801–11.

Kretzschmar, M., Reinking, D. P., Jager, J. C., Brouwers, H., van Zessen, G. (1994). 'Network models: from paradigm to mathematical tool', in Kaplan, E. H., Brandeau, M. (eds) *Modeling the AIDS epidemic*. New York: Raven Press, 561–83.

Longini Jr., I. M., Clark, W. S., Satten, G. A., Byers, R. H., Karon, J. M. (1996). 'Staged Markov models based on CD4+ T-lymphocytes for the natural history of HIV infection', in Isham, V., Medley, G. *Models for infectious human diseases: their structure and relation to data*. Cambridge: Cambridge University Press, 439–59.

Longini, Jr., I. M., Clark, W. S., Byers, R. H., Ward, J. W., Darrow, W. W., Lemp, G. F., Hethcote, H. W. (1989a). 'Statistical analysis of the stages of HIV infection using a Markov model'. *Statistics in Medicine* 8: 831–43.

Longini Jr., I. M., Clark, W. S., Haber, M., Horsburgh Jr., R. (1989b). 'The stages of HIV infection: waiting times and infection transmission probabilities', in Castillo-Chavez (ed.) *Mathematical and statistical approaches to AIDS epidemiology*, 111–37.

McLean, A. R., Blumberg, B. S. (1994). 'Modelling the impact of mass vaccination against hepatitis B. I. Model formulation and parameter estimation'. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 256: 7–15.

Morris, M., Kretzschmar, M. (1997). 'Concurrent partnerships and the spread of HIV'. *AIDS* 11: 641–8.

Papaevangelou, G., Richardson, S. C. (1995). 'HIV prevalence and risk factors among injecting drug users in EC and COST countries', in Baert, A.-E., et al. *AIDS research at EC level*. Amsterdam: IOS Press, 73–82.

Peterson, D., Willard, K., Altmann, M., Gatewood, L., Davidson, G. (1990). 'Monte Carlo simulation of HIV infection in an intravenous drug user community'. *Journal of AIDS* 3: 1086–95.

Richardson, C., Ancelle-Park, R., Papaevangelou, G. (1993). 'Factors associated with HIV seropositivity in European injecting drug users'. *AIDS* 7: 1485–91.

Williams, J. R., Nokes, D. J., Medley, G. F., Anderson, R. M. (1996a). 'The transmission dynamics of hepatitis B in the UK: a mathematical model for evaluating costs and effectiveness of immunization programmes'. *Epidemiology and Infection* 116: 71–89.

Williams, J. R., Nokes, D. J., Anderson, R. M. (1996b). 'Targeted hepatitis B vaccination – a cost-effective immunisation strategy for the UK?' *Journal of Epidemiology and Community Health* 50: 667–73.

Yakowitz, S. J. (1994). 'From a microcosmic IDU model to a macrocosmic HIV epidemic', in Kaplan, Brandeau, *Modeling the AIDS epidemic*, 365–83.

CHAPTER 11

DEVELOPING MULTINATIONAL SCENARIO ANALYSES OF HEALTH IMPACTS OF DRUG USE

Johannes C. Jager, Maarten J. Postma and Peter W. Achterberg

The issue of drug use and its health-related consequences represents a complicated and multifarious scientific and societal problem. This chapter explores how scenario analysis may be used to investigate the actual situation, possible future developments and potential effects of policy measures in Europe.

Since the early 1980s, considerable experience with future research on public health and health-care problems has been built up in the Netherlands. The Dutch Steering Committee on Future Health Scenarios (STG) commissioned about 20 scenario studies on diseases and their consequences for public-health-strategy development, including on chronic diseases such as diabetes mellitus, chronic non-specific lung disease, rheumatoid arthritis (Casparie and Verkleij, 1991–94), mental health care (Idenburg et al., 1991; Bijl, 1991) and HIV/AIDS (van den Boom et al., 1992). The methodological accomplishments from these STG studies have been summarised and translated into practical guidelines for scenario research (Genugten et al., 1996). On a broader scale, covering many diseases and determinants (risk factors) which affect health, the Department for Public Health Forecasting of the National Institute of Public Health and the Environment (RIVM), Bilthoven, established, in cooperation with several other institutes, a detailed description including forecasts of the health status of the Dutch population (Ruwaard et al., 1994; Ruwaard and Kramers, 1997). Further, in two specific research projects, the scenario-analytic approach was extended to the multinational (European Union) level, with the European Commission Concerted action on multinational AIDS scenarios examining the epidemiological, economic and socio-cultural impact of AIDS on society (Jager et al., 1995; Downs et al., 1997; Postma et al., 1997) and another project on the attainment of blood self-sufficiency in the EU (Van Aken et al., 1997).

These multinational EU-directed studies have developed and implemented an integrated application of mathematical modelling, epidemiology and health economics within a scenario-analytic framework. This chapter builds on the experience acquired through the studies with respect to their relevance for the scenario research process. The second section characterises essential features of the scenario-analytic approach while the third section motivates the application of scenario analysis to the drug issue. The following section gives a concise overview of the multinational scenario projects mentioned above

and their generalisable methodological accomplishments, and the fifth section offers a preliminary approach to scenario-building in modelling drug use and related problems. The last section contains some conclusions and recommendations for potential policy-oriented research on drug use and related problems.

The scenario-analytic approach

Scenario analysis is a form of research that makes use of a variety of techniques, such as mathematical modelling and statistical analysis, brain-storming techniques, meta-analysis and cost-effectiveness analysis. Its main power lies in structuring a joint study involving participants from different disciplines, interests and countries. It differs from empirical research in its strong dependence on secondary data (data not collected according to a strict study design and sampling plan), incomplete and heterogeneous data and its meta-analytic approach. Methodological weakness has to be remedied by extensive consultation of experts.

Essential elements of a scenario research are the conceptual approach, the use of mathematical models, departure from a basic analysis and the development of scenarios.

Conceptual model

A scenario study starts with the construction of a conceptual model. This construction should be a joint activity of all participants involved in the project. The conceptual model is a schematic representation of the field under study in terms of components and their main relations. As stated in Ruwaard et al. (1994), it has the following functions: to provide a structure for the development of ideas; to define the boundaries; to offer a structure for orderly handling of the subject matter; and to serve as a basis for producing a formal model in mathematical terms.

Mathematical model

Ideally, the conceptual model is elaborated as a mathematical model, which can be used to describe the past, to predict the future and to perform 'what if' analyses. Modelling results represent the quantitative backbone of the scenario development. The role of modelling and the interaction of modellers and other researchers has been discussed in Jager and van den Boom (1994). For a discussion of the interdependencies of quantitative and qualitative procedures in scenario research, see Bijl (1991).

Basic analysis

The basic analysis (or baseline analysis) comprises the identification of data sources and the collection of available, primary and secondary data on all the components taken up in the conceptual model. These are subject to data analysis and meta-analysis, and form the basis for the application of scenarios. The evaluation of data availability and quality (comparability, completeness, potential for generalisation) is important here. Another

important aspect of scenario analysis is to specify data needs. The basic analysis also comprises the selection and specification of key variables and impact indicators to be handled in the construction of scenarios (and the mathematical model). The data play a role in the validation of models. Often modelling detects essential data gaps. Key variables represent factors which may significantly influence the system under study. Impact indicators describe the most interesting aspects of the performance of the system.

Scenario development

The construction of scenarios starts from the conceptual model, the mathematical model (where possible) and the results of the basic analysis. A reference scenario builds on the assumption that the present situation and recent trends will continue in the future, and describes the surprise-free continuation of the past. Alternative scenarios may be elaborated on the basis of 'what if' questions.

Why apply a multinational scenario approach to drugs?

Generally, the scenario-analytic approach to problems within the domains of public health and health care is motivated by its expected contribution to policy-making, to long-term planning and to strategy development. Effective analyses supporting these activities require a methodology which deals coherently with epidemiological analyses of the current situation and its projection into the future based on explicit assumptions. Mathematical modelling techniques and linking epidemiology to insights in costs related to care and prevention from health-services research are essential (Jager and van den Boom, 1994).

Thus, a scenario study requires cooperation between specific experts such as modellers, epidemiologists and economists. The following issues will particularly influence the choice of a scenario approach:

- the weight, scope and complexity of the problem (only important and complex problems require a scenario approach; there is a demand for interdisciplinary approach);
- the need to make decisions and launch interventions, which may be confronted with great uncertainty (an interest in the future; the wish to make predictions and plans/interventions, which are extremely difficult in view of great uncertainty);
- the potential impact on decision-making by elucidating alternative developments in quantitative terms;
- the involvement of different groups (patients, care providers, public-health officials, community, government, countries) with diverging interests and the inherent invitation to discuss these alternatives in a coherent and rational way (potential consensus as a result of discussion between interested parties and relevant disciplines).

To conduct a successful scenario analysis it should be possible to distinguish key variables and impact indicators that are sensitive to policies.

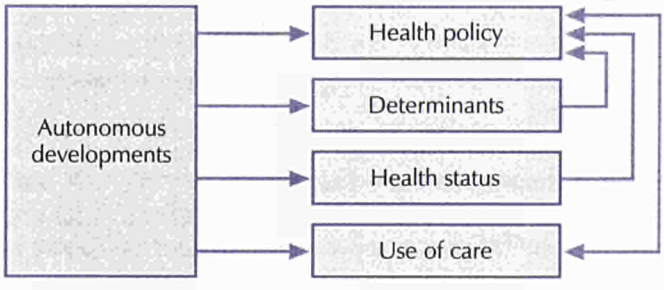
It is easy to show that the drug issue fulfils the above characteristics of a problem which should be investigated with the help of scenario analysis. The drug problem is worldwide, it depends on widely different socio-cultural conditions/factors, it is being addressed by different policies in different countries, it is in urgent need of effective interventions and is a 'hot topic' politically. It involves drug users, their partners and families, the health and legal systems, health and various other officials and it causes discussions between countries and between disciplines such as medicine, epidemiology, the economy and law, the study of social problems including ethics and the methodological disciplines (mathematics and statistics).

Experiences from multinational and multidisciplinary scenario studies

Conceptual modelling

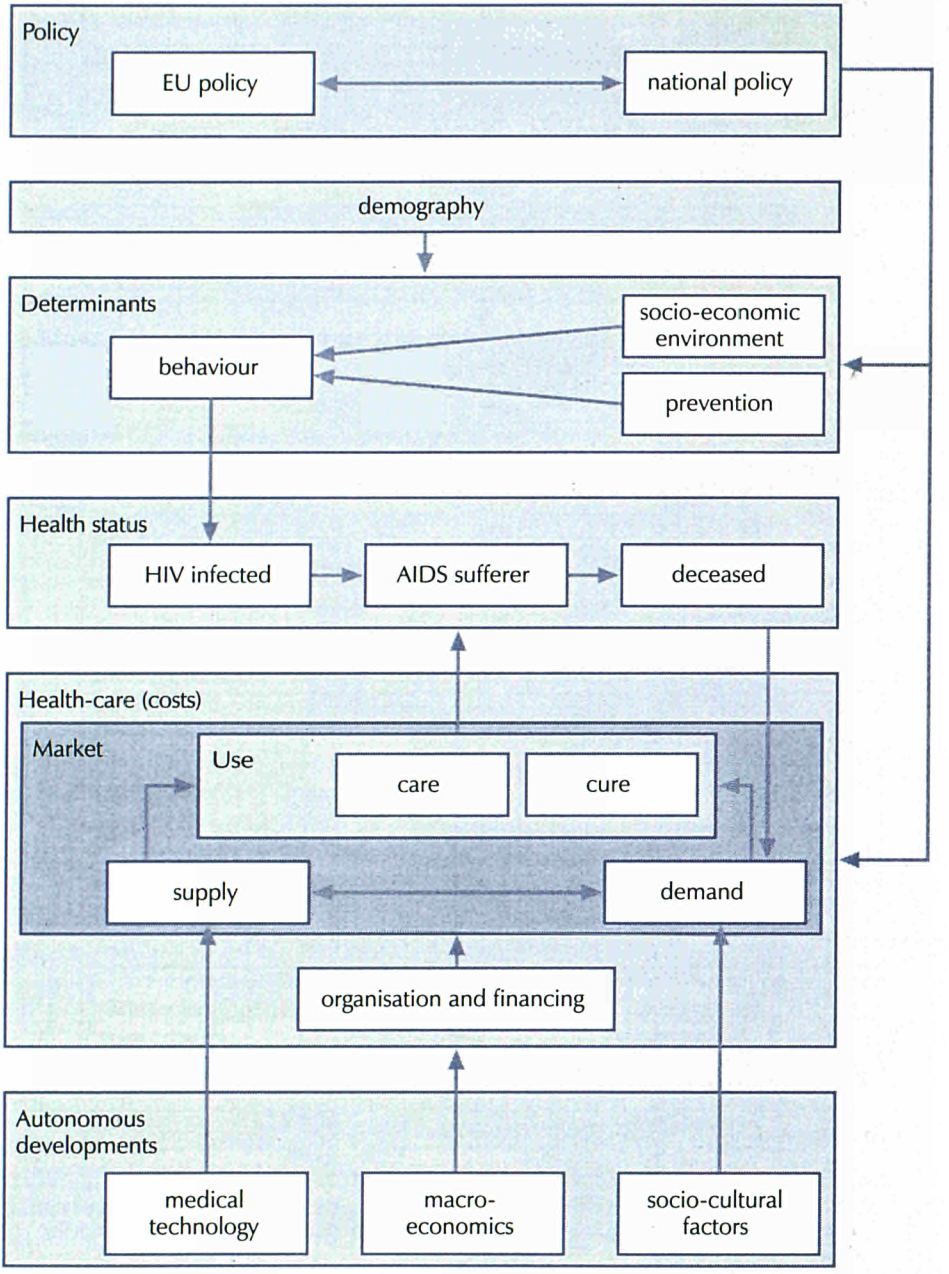
A conceptual model designed for public-health-status assessment and forecasts comprises five basic components: health status; determinants of health status; health policy; use of care (and costs) and autonomous developments; and their interrelationships (Ruwaard and Kramers, 1997) (see Figure 1). For further elaborations of the separate components, see Ruwaard et al. (1994) and Ruwaard and Kramers (1997). Specific health problems which have already been analysed have been modelled in an analogous way. The conceptual model underlying the EU Concerted action on multinational AIDS scenarios (Jager et al., 1995) is presented in Figure 2. The scenario study on the attainment of blood self-sufficiency in the EU (van Aken et al., 1997) has been based on a conceptual model of a national blood-supply system, incorporating policies, population characteristics, donors, the production methods of blood products, the product-specific market, the recipients and the medical methods and performance (see Figure 3). In all studies, the construction of the conceptual scheme stimulated enthusiastic discussions and resulted in a clear delineation of the subject under study, the specification of relations which should be studied and corresponding appointments for the division of tasks among the research teams involved and their prioritisation. In this way, conceptual

Figure 1: Conceptual basic model of public-health status and forecasts, 1997



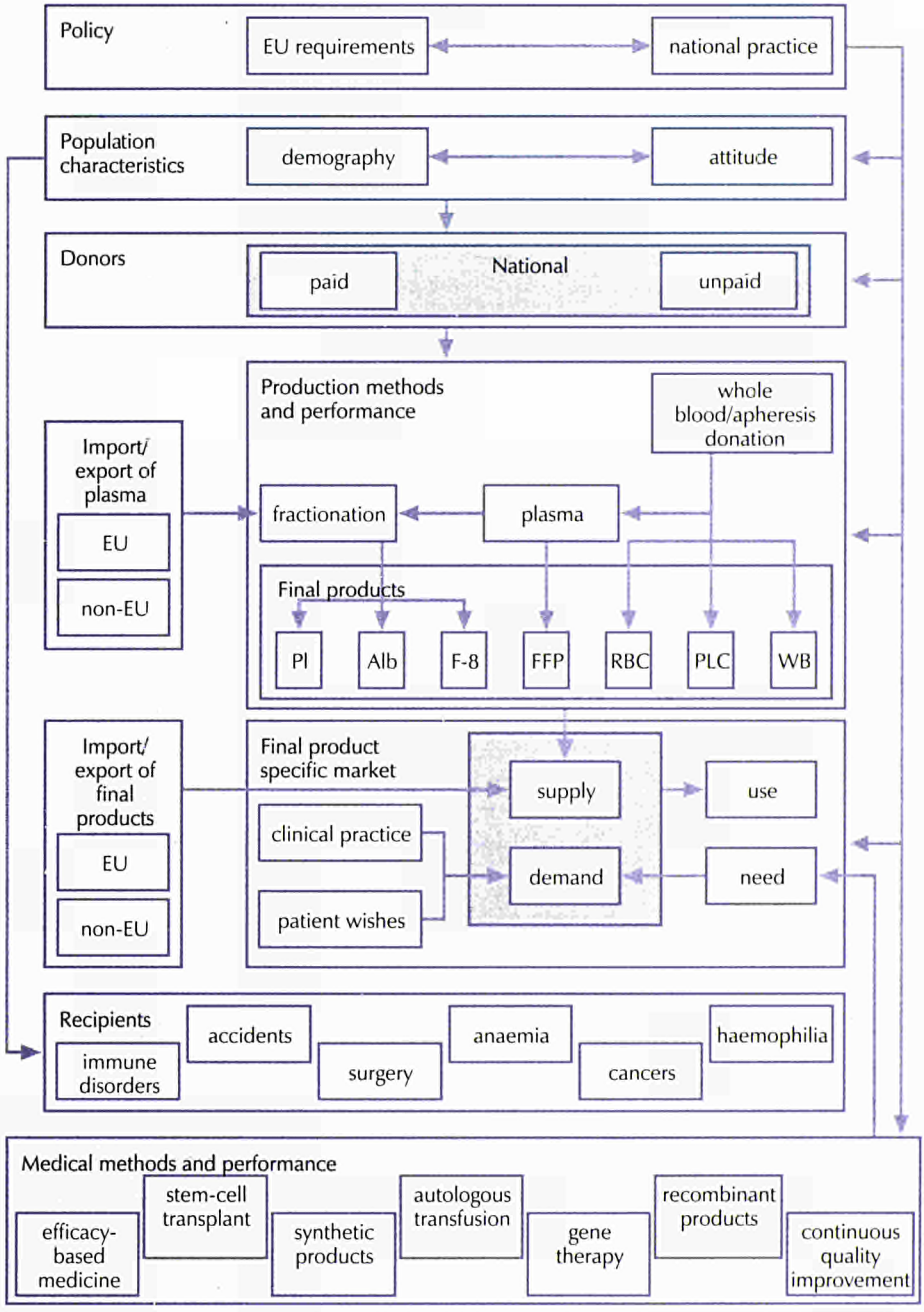
Source: Ruwaard and Kramers (1997)

Figure 2: Conceptual model for the European Commission Concerted action on multinational AIDS scenarios



Source: Jager et al. (1995)

Figure 3: Conceptual model for national blood supply



Source: van Aken et al. (1997)

modelling stimulates and facilitates interdisciplinary discussion and provides the mathematicians with starting points for elaborating the conceptual model in mathematical terms. This process can also highlight specific data needs.

Mathematical modelling

Mathematical modelling provides several functions: quantitative description of the phenomenon under study; insight into the causal relations involved; detection of crucial data needs; parameter estimation starting from empirical information; projection into the future under specified conditions; and prediction. In the study concerning AIDS-impact scenarios, several models have been developed which pertain to parts of the conceptual model (e.g. concerning the effects of medical treatment, behavioural change and the organisation of health-care provisions) (Bailey and Heisterkamp, 1995). It would be unrealistic to try to encompass the whole area under study in one, necessarily very complicated, mathematical model. However, the specifications for different models were similar enough to warrant comparability of the outcomes. Much attention was paid to the development and application of a common methodology to describe the baseline situation (Downs et al., 1997; Heisterkamp, 1995). The blood-supply problem (van Aken et al., 1997) has been represented by a single model with sub-systems which may be activated as desired.

The overall experience is that mathematical modelling is essential in scenario analysis, but it should be applied flexibly (several models using the same specifications; simultaneous use of simple descriptive (trend analyses) and more complex dynamic models) and results should be supplemented by qualitative considerations, discussions and comments.

In AIDS-scenario studies, several models have been developed which incorporate intravenous drug use (IDU) as a determinant in the spread of HIV. For the Dutch AIDS scenario study (van den Boom et al., 1992) a multi-group transmission model was developed (van Druten et al., 1990, 1992) which represents the HIV epidemic as resulting from overlapping sub-epidemics in six groups (among which are IDU men and women). The contacts between and within these groups occur via several transmission routes: unprotected anal intercourse; unprotected vaginal intercourse; and the shared use of needles. The model enables the simulation of the effects of preventive interventions targeted to behavioural change (see Chapter 7 in van den Boom et al., 1992). Another mathematical technique, which also concerns the spread of HIV in several risk groups, including IDUs, is the so-called back calculation (BC) (Heisterkamp, 1995, and Chapter 6, above). Back calculation makes it possible to estimate past and to project future incidence and prevalence of HIV infection based on reported AIDS surveillance data. In the EC project on multinational AIDS scenarios, these estimates were obtained for the 15 European Member States (Downs et al., 1997). Recently, this BC has been extended to include age effects (report of EC Concerted action on multinational AIDS scenarios, in preparation). This advancement is important because it supports the results of analysis of the AIDS epidemic in birth cohorts of separate risk groups such as IDUs (Houweling et al., 1998) for public-health-strategy development. Recent estimates of HIV infections in IDUs in Europe by BC are available.

Basic analysis (key variables and impact indicators)

Basic analysis involves the collection of available data, which can be specified according to variables associated with the components in the conceptual model. It has been found profitable to devise a table which presents relevant variables according to the components of the conceptual model and/or according to the disciplines involved (van den Boom et al., 1992; Jager et al., 1995; van Aken et al., 1997). The correspondence to related impact indicators should be given. This table concerning the AIDS-scenario study is structured according to disciplines such as epidemiology, economy and sociology (see Figure 4) (Jager et al., 1995). In the blood-supply study (van Aken et al., 1997) the impact indicators were explicitly developed by considering each component in the model (see Figure 5).

Scenario development

In most previous scenario studies (van den Boom et al., 1992; Jager et al., 1995; van Aken et al., 1997), the construction of the reference scenario generally took a major part of the available time. In an initial planning meeting for the project on multinational AIDS scenarios, four working groups were formed on modelling support, on epidemiological impact, on economic impact and on social impact, respectively. The mathematicians started to devise an overview of available mathematical models (Heisterkamp, 1995) and decided to develop a common methodology which could be applied to European AIDS surveillance data. The epidemiologists needed several sessions to specify – in cooperation with the modellers – the epidemiological starting points for the reference scenario (Houweling et al., 1996). They also had to collect additional country-specific HIV/AIDS surveillance data to underpin their specifications (Roosmalen et al., 1996). Thus, rather late during the course of the project, alternatives to the reference scenario were eventually specified, though in principle, at the beginning of the project, available mathematical models could handle them very well. This shows that a reference scenario should be developed as soon as possible in the execution of a scenario project.

Apparently, a detailed elaboration of a reference is a satisfactory foundation for formulating themes for alternative scenarios and further scenario-building. Another aspect of multidisciplinary scenario construction refers to levels, for example the extension of the epidemiological level to the economic level by linking epidemiological variables to economic information (see Figure 4, and Leidl, 1994). The mathematicians should keep in mind which level is an endpoint (in terms of impact indicators) for their modelling activity, because it will determine the set-up of their models.

Some notes on economic aspects

Classification of costs and definition of economic-impact indicators

Essential dichotomies regarding the classification of costs suitable for scenario analyses of costs and effects are: 'costs within the health-care system' versus 'costs outside the health-care system'; and 'direct costs' versus 'indirect costs' (van Genugten et al., 1996, and Chapter 12, below).

Figure 4: Key variables and impact indicators (epidemiological and economic level) for the EC Concerted action on multinational AIDS scenarios

Key variables	Impact indicators
Epidemiological level	
Demography	Morbidity: Incidence, Prevalence, Person-years
Age	
Gender	
Country	Mortality
Time (time scale of effects)	
AIDS definition	
Progression of disease	
Natural history	
incubation time	
Medical history/intervention	
pre-AIDS/AIDS treatment	
disease stage/survival	
(AIDS as chronic disease)	
new therapies	
vaccination	
Risk group	
Risk factor	
Economic level	
Per person-year health-care needs/costs	Direct costs: Patient-related: Hospital beds/outpatient contacts/care costs/costs of treatment: Medication: costs; Nursing home: places/costs; GP: contacts/costs; Home care: contacts/costs; Other outpatient: contacts/costs; General programme; Prevention; Research
	Indirect costs: Potential years of life lost
Organisation of care	

Source: Jager et al. (1995)

Figure 5: Key variables and impact indicators for blood self-sufficiency in the European Community, classified according to the components of the conceptual model

Component	Autonomous developments	Key variables	Impact indicators
EU requirements National policy		safety regulations	% unpaid donations/donors
Population	demography		
Donors	new virus	donor attitudes no. of unpaid donors no. of paid donors no. of donations % of whole blood donors % of plasmapheresis donors % of apheresis donors quality of donations volume of imported plasma for fractionation amount of imported surplus red-blood-cell concentrates	need demand supply shortage use
Production	legal regulations	safety regulations yields from donations yields of products quarantine period improved storage and transport platelet technology with less plasma platelet apheresis versus whole blood donation	% self-sufficiency at national and/or EC level % import % export
Market	financial implications	supply of alternatives clinical practices wastage	
Recipients	trends in accidents trends in diseases trends in surgery		
Medical technology		use of recombinant products use of synthetic products % of autologous transfusions % of stem-cell transplantations clinical standards clinical control introduction of gene therapy	

Source: van Aken et al. (1997)

Linking epidemiology and health economy

In linking epidemiology and health economy the following notions are important for constructing scenarios concerning the economic consequences of health problems: 'incidence-based indicators' versus 'prevalence-based indicators'; the connection of these indicators to successive economically relevant disease stages or longitudinal profiles of care (Dijkgraaf et al., 1996; Postma et al., 1995); and resource use 'per patient' versus resource use 'per person-year'. These issues are illustrated in detail in Chapter 12 (see also Scitovsky and Over, 1988).

Scenario analysis versus cost-effectiveness analysis

There is great similarity between guidelines for scenario analysis and for the conduct of cost-effectiveness analyses (CEA) for the allocation of health-care resources. Recent guidelines for CEA (Russell et al., 1996) stress the importance of a reference case with a broad perspective, as a point of comparison for specific alternatives. In addition, the crucial role of conceptual modelling and its relation to mathematical modelling are recognised (Gold et al., 1996). Thus, on the one hand, any CEA should include structured scenarios and, on the other hand, CEAs can easily be incorporated into scenario analyses. CEAs and scenario analyses can thus be mutually linked.

Conceptual model for drug use and health-related consequences

The basic conceptual model underlying the Dutch public-health status and forecasts effort in 1997 (Ruwaard and Kramers, 1997) is given in Figure 1, above. It comprises five coherent components or blocks: health policy; determinants; health status; use of care; and autonomous developments. Health status takes a central position and is affected by the determinants of health, which in turn can be affected by health policy. The health status of the population leads to the use of care. Health policy is designed in response to developments concerning determinants, health status and use of care. Autonomous developments affect this system. The arrows represent the main interrelations of interest. The determinants comprise somatic and mental health care and different components of prevention such as health protection, disease prevention and health promotion, which together with exogenous factors (physical environment, lifestyle, social environment) and endogenous (hereditary, acquired) factors influence health status. Departing from such representations as developed in Ruwaard et al. (1994) and Ruwaard and Kramers (1997), a preliminary conceptual model for the health-related consequences of drug use in the EU has been drafted (Figure 6).

Table 1 lists suggested components for a preliminary conceptual model for drug use and (health-related) consequences. Figure 6 shows the complicated position of drug use as an element of lifestyle affecting health and affected by many other factors as indicated by the arrows in the figure. Here drug use is placed within the context of public health. The conceptual model can be adapted to other relevant contexts. A study focusing on the impact of one determinant or on a specific health condition should start with the quantitative assessment of (trends in) incidence, prevalence and (geographic) spread, as indicated in the table. Thus the quantitative measurement of the impact of drug use starts with the assessment of incidence, prevalence, trends and geographic spread of drug use.

Figure 6: Preliminary conceptual model for drug use and health-related consequences

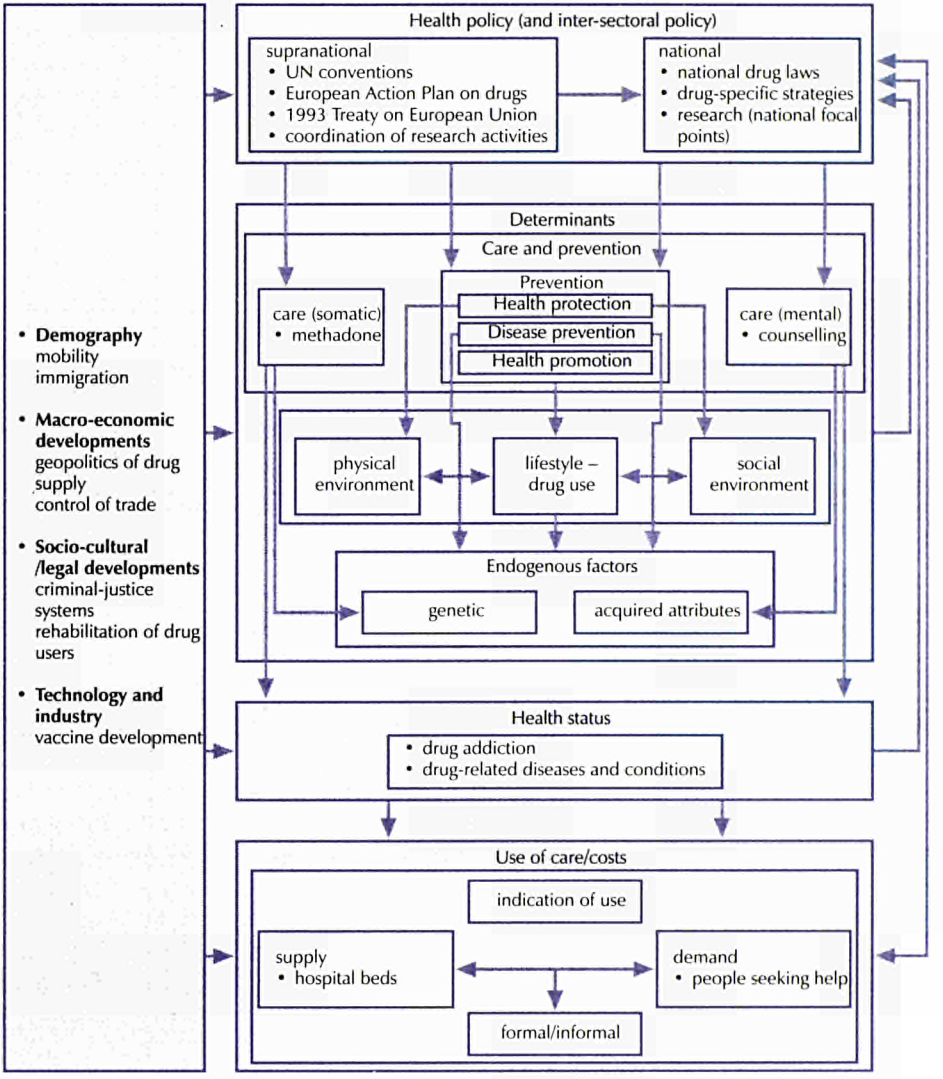


Table 1: Components for conceptual modelling of drug use and its consequences

- Health policy (and intersectoral policy)
  - Supranational strategies/policies
    - United Nations conventions and organisations (WHO)
    - European Action Plan on Drugs (EU)
    - Treaty on European Union 1993
    - Coordination of research activities
    - National strategies/policies
    - Drug-specific strategies
    - Research (national focal points)
- Determinants (risk factors) affecting health
  - Health care and prevention
    - Somatic care
      - Substitute prescribing (methadone)
      - Detoxification
      - Pharmaceuticals
    - Prevention
      - Health protection
        - Syringe and needle exchange
      - Disease prevention
        - Vaccination
        - Screening for drug-related infectious diseases
      - Health promotion
        - Education (safe sex/needle hygiene)
        - Targeted programmes
        - Professional training
        - Mass-media campaigns
        - Telephone help-lines
      - Mental care
        - Counselling
  - Exogenous factors
    - Physical environment
      - Biological agents
        - HIV/HVB/HVC/*M. tuberculosis*
        - Resistance to antibiotics
      - Physical factors
        - Housing
      - Chemical factors
        - Quality of drugs
    - Lifestyle (quantitative assessment of incidence/prevalence/trends/spread)

continued

- Drug use
  - Type of use
    - Hard/soft drugs
    - Route of administration (by mouth, smoking, injecting, sniffed)
    - Needle sharing
    - Frequency/prevalence/patterns of use
  - Sexual behaviour/prostitution
  - Alcohol
  - Lack of hygiene/dental care
- Social environment
  - Sub-populations (sub-cultures/social profiles of users)
    - General population
    - Ethnic minorities
    - School children
    - Young people outside school
    - Students and intellectual circles
    - Geographic patterns/urbanisation
    - Availability and supply of drugs
  - Attitudes/public opinion

Endogenous factors

- Genetic factors
  - Gender differences
- Acquired attributes
  - Ageing
  - Immunodeficiency
  - Malnutrition

Health status (quantitative assessment of health impact: incidence, prevalence of disease/disease stages, mortality, trends, spread, transmission)

- Drug addiction (dependence, poly-drug dependence)/overdose/accidents/suicides
- Drug-related diseases and conditions
  - HIV/hepatitis (B/C)/endocarditis/tuberculosis
  - Injection-related abscesses/malnutrition/weight loss/respiratory problems/nasal damage/sleep disturbance/chest pain
  - Neurological/psychiatric disorders
  - Comorbidity

Use of care/costs

- Formal/informal care
- Need
  - People needing help
  - Objective versus subjective need
  - Denial

continued

- Supply
  - Hospital beds
  - Budget
  - Availability of doctors
- Demand
  - People seeking help
  - Supplier-induced demand
  - (Un)satisfied
- Indication/use
  - Hospital
  - GP
  - Medication

Related costs

- Direct costs
  - Patient-related
  - General programme
- Indirect costs
  - Lost productivity
  - Employment
  - Potential years of life lost

Autonomous developments

- Demography
  - Mobility/immigration
- Macro-economic developments
  - Geopolitics of supply/control of trade
- Socio-cultural/legal developments
  - Criminal-justice systems/national drug laws (legal versus illegal drugs)
  - Accessibility (prohibiting/punishing possession/arrests and drug seizures)
  - Reduction in demand/supply of drugs
    - Rehabilitation of drug users
- Medical technology
  - Vaccine development

Conclusions

Scenario analysis offers an appropriate methodological context for approaching the drug problem in a policy-oriented multinational sphere of interest. It can stimulate various research efforts in important interdisciplinary areas related to the drug-use problem in the EU Member States, so as to allow for best practices and best research to disseminate throughout the EU. This will guide individual States towards optimal strategies for combating the drug problem.

Four (general) aspects of scenario analysis should be emphasised: conceptual modelling; its relation to data-gathering (basic analysis); its relation to mathematical modelling; and its potential for orderly and creative scenario construction.

Conceptual modelling stimulates and facilitates interdisciplinary discussions and provides mathematicians with starting points for elaborating the conceptual model in mathematical terms. A multinational approach will allow for a much broader and divergent input to conceptual modelling, with a greater chance of high-quality expert input in all relevant areas. The process of conceptualisation stimulates causative and associative thinking and may identify priorities for data collection and analysis.

The drug problem should be delineated in a conceptual model. A preliminary conceptual model has been suggested as a starting point for further discussion on refinements and further elaboration. Parts of the conceptual model and associated specific questions may be selected for mathematical modelling. The development of scenarios should also build on a conceptual model.

Mathematical modelling should be applied in a flexible way within a scenario analysis, for example, by using several models simultaneously, such as simple, descriptive models and more complex, dynamic models. Modelling results should be supplemented by qualitative considerations to guide further discussions and comments, possibly resulting in new and better model input. Again, the multinational approach will give access to increasingly diverse mathematical models and expertise. Mathematical models should be classified in relation to the components and relations indicated in the conceptual model. The potential of mathematical models with respect to the construction of a reference scenario and alternative scenarios should be made explicit.

There are several other advantages of, and opportunities to be gained from, a multinational approach. Assessing variations within and between countries broadens the view of different policy options. A multinational approach may provide what might be called a 'natural' set of experiments as a disease may express itself differently in different countries. It can thereby also provide a first realistic estimate of the variability of certain outcomes. A good example, again, is the AIDS epidemic, where the development of an epidemic in a particular country is highly dependent upon the magnitude and characteristics of the sub-epidemics.

A multinational approach will thus give rise to an optimally balanced scenario analysis with regard to expertise in the areas relevant to the analysis, that is, epidemiology, mathematical modelling, economic analysis, and medical and sociological expertise. Moreover, policy-makers from different countries may in certain stages also be involved and contribute to the variety in policy options.

Acknowledgements

This chapter benefited from the support of the European Commission's Concerted action on multinational AIDS scenarios (BMH1-CT-941723) and the Scenario analysis regarding the attainment of blood self-sufficiency in the European Community (SOC94 200985 05F01(94CVVF1-040-0). The authors would like to thank G. J. Kommer, W. G. Martens and F. J. Prinsze for their help in drawing the figures.

References

van Aken, W. G., Jager, J. C. (project leaders), Kommer, G. J., Magirou, V. F., Postma, M. J., Prastacos, G. P., Strengers, P. F. W. (research team), Delaney, F. M. (editor), Delaney, F. M., McClelland, D. B. (advisors) (1997). *Scenario analysis regarding the attainment of blood self-sufficiency in the European Community*. A report for the European Commission (SOC94 200985 05F01 (94CVVF1-040-0)). Amsterdam/Bilthoven/Athens: CLB, RIVM, OMAS.

Bailey, N. T. J., Heisterkamp, S. H. (eds) (1995). *Preliminary review on the mathematical models and data used for the baseline analysis*. Bilthoven: National Institute of Public Health and the Environment (RIVM).

Bijl, R. V. (1991). 'Construction of future scenarios: policy-oriented scenario research, applied to the field of mental health and mental health care'. Thesis, Netherlands Centre for Mental Health, Utrecht.

van den Boom, F. M. L. G., Jager, J. C., Reinking, D. P., Postma, M. J., Albers, C. E. S. (1992). *AIDS up to the year 2000. Epidemiological, sociocultural and economic analysis for the Netherlands*. Scenario report commissioned by the Steering Committee on Future Health Scenarios. Dordrecht/Boston/London: Kluwer Academic Publishers.

Casparie, A. F., Verkleij, H. (1991–94). *Chronic diseases in the year 2005*. Scenario reports commissioned by the Steering Committee on Future Health Scenarios. Dordrecht/Boston/London: Kluwer Academic Publishers. Volume 1: *Scenarios on diabetes mellitus 1990–2005*, 1991. Volume 2: *Scenarios on chronic non-specific lung diseases 1990–2005*, 1993. Volume 3: *Scenarios on rheumatoid arthritis 1990–2005*, 1994.

Dijkgraaf, M. G. W., Luijben, A. H. P., Postma, M. J., Borleffs, J. C. C., Schrijvers, A. J. P., Jager, J. C. (1996). 'Lifetime hospitalization profiles for symptomatic, HIV-infected persons'. *Health Policy* 35: 13–32.

Downs, A. M., Heisterkamp, S. H., Brunet, J.-B., Hamers, F. F. (1997). '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* 11: 649–62.

van Druten, J. A. M., Poos, M. J. J. C., Hendriks, J., Jager, J. C. (1992). 'HIV epidemics in linked risk groups: a move to exploratory scenario analysis', in Jager, J. C., Ruitenberg, E. J. (eds) *AIDS impact assessment: modelling and scenario analysis*. Amsterdam/London/New York/Tokyo: Elsevier, 83–94.

van Druten, J. A. M., Reintjes, A. G. M., Jager, J. C., Heisterkamp, S. H., Poos, M. J. J. C., Coutinho, R. A., Dijkgraaf, M. G. W., Ruitenberg, E. J. (1990). 'HIV infection dynamics and intervention experiments in linked risk groups'. *Statistics in Medicine* 9: 721–36.

van Genugten, M. L. L., Rutten, F. F. H., Jager, J. C. (1996). *Scenario development and costing in health care. Methodological accomplishments and practical guidelines*. Report commissioned by the Foundation for Future Health Scenarios. Utrecht: International Books.

Gold, M. R., Siegel, J. E., Russell, L. B., Weinstein, M. C. (1996). *Cost-effectiveness in health and medicine*. New York/Oxford: Oxford University Press.

Heisterkamp, S. H. (1995). *Quantitative analysis of AIDS/HIV: development of methods to support policy making for infectious disease control*. Utrecht: Elinkwijk BV.

Houweling, H., Hamers, F., Termorshuizen, F., Gill, O. N., Jager, J. C., Coutinho, R. A. (1998). 'A birth cohort analysis of AIDS in Europe: high incidence among young persons at risk'. *AIDS* 12: 85–93.

Houweling, H., Heisterkamp, S. H., Gill, O. N., Jager, J. C., Coutinho, R. A. (1996). *Specification of the epidemiological reference scenario*. Bilthoven: RIVM.

Idenburg, P. A., Ketting, E., Bijl, R. V., et al. (1991). *Caring for mental health in the future. Future scenarios on mental health and mental health care 1990–2010*. Dordrecht/Boston/London: Kluwer Academic Publishers.

Jager, J. C., van den Boom, F. M. L. G. (1994). 'Scenario analysis, health policy, and decision-making', in Kaplan, E. H., Brandeau, M. L. (eds) *Modelling the AIDS epidemic*. New York: Raven Press, 237–52.

Jager, J. C., Postma, M. J., Tolley, K., Kennelly, J. (1995). 'The economic impact of AIDS: from national towards multinational scenarios', in Fitzsimons, D., Hardy, V., Tolley, K. (eds) *The economic and social impact of AIDS in Europe*. London: Cassell, 40–53.

Leidl, R. M. (1994). 'Some factors to consider when using the results of economic evaluation studies at the population level'. *International Journal of Technology Assessment* 10: 467–78.

Postma, M. J., Jager, J. C., Dijkgraaf, M. G. W., Borleffs, J. C. C., Tolley, K., Leidl, R. M. (1995). 'AIDS scenarios for the Netherlands: the economic impact on hospitals'. *Health Policy* 31: 127–50.

Postma, M. J., Tolley, K., Leidl, R. M., Downs, A. M., Beck, E. J., Tramarin, A. M., Flori, Y. A., Santin, M., Antoñanzas, F., Kornarou, H., Paparizos, V., Dijkgraaf, M. G. W., Borleffs, J. C. C., Luijben, A. H. P., Jager, J. C., for the European research team on AIDS scenarios (1997). 'Hospital care for persons with AIDS in the European Union'. *Health Policy* 41, 157–76.

van Roosmalen, M., Hamers, F., Downs, A. M., Houweling, H. (1996). *Results of the questionnaire on country-specific HIV/AIDS surveillance data to be used in back-calculation*. Bilthoven: RIVM.

Russell, L. B., Gold, M. R., Siegel, J. E., Daniels, N., Weinstein, M. C. (1996). 'The role of cost-effectiveness analysis in health and medicine'. *Journal of the American Medical Association* 76 (14): 1172–7.

Ruwaard, D., Kramers, P. G. N. (eds) (1997). *Volksgezondheid toekomst verkenning 1997. De som der delen*. Amsterdam: Elsevier/De Tijdstroom.

Ruwaard, D., Kramers, P. G. N., van den Berg Jeths, A., Achterberg, P. W. (1994). *Public health status and forecasts. The health status of the Dutch population over the period 1950–2010*. The Hague: Sdu Uitgeverij.

Scitovsky, A. A., Over, M. (1988). 'AIDS: costs of care in the developed and the developing world'. *AIDS* 2 (suppl. 1): S71–S81.

PART V

ECONOMIC MODELS

Introduction

Economic models are often directly translatable to decisions on budget allocation, which is of primary importance for policy-makers. In Chapter 12, Postma and his colleagues elaborate on estimating drug-related HIV/AIDS costs as an example for extending to other drug-related health-care costs. HIV/AIDS health-care costs include costs for resources such as hospital beds, nursing care, specialist consultations, day care and GP visits. Costing studies have been used to assess and project national health-care use and costs for HIV/AIDS and the economic impact of different intervention scenarios has also been explored. A set of methods thus exists that can be applied to other drug-related diseases. To date, studies have used different methods, and the authors argue that greater standardisation is necessary. They recommend extending the work on HIV/AIDS to hepatitis as a first step, collecting country-specific information needs on health-care use and costs by drug users, and developing future costing scenarios for all relevant drug-related diseases.

Sutton describes the potential for modelling economic drug markets in Chapter 13. He distinguishes three levels of analysis: picoeconomics (individual level); microeconomics (areas or markets); and macroeconomics (across areas or markets). Studies at individual level have the greatest potential, but the macro level is what policy-makers are most interested in and where by far most data are available. The author suggests that price and income elasticities of demand are currently unclear from existing published studies, and that essential issues to explore are individual health perception and education, individual preferences, habit formation, time preference and peer effects. Sutton concludes that individual-level analysis by a multi-disciplinary team is probably the most promising modelling approach for understanding interactions between policy interventions and markets.

CHAPTER 12

MODELLING THE HEALTH-CARE COSTS OF DRUG-USE-RELATED DISEASE

Maarten J. Postma, Keith Tolley and Johannes C. Jager

Different types of costs, such as health-care costs, the social costs of dependency and costs associated with criminal activity, can be related to drug use. In general, drug users are at increased risk of various types of infectious diseases, such as HIV/AIDS and hepatitis (Haverkos and Lange, 1990). Furthermore, some non-infectious diseases have also been associated with drug use, in particular with the use of crack cocaine (Umbricht-Schneiter et al.,1994; Kaku and Lowenstein, 1990; Dinwiddie et al., 1992). An overview of drug-related diseases is provided in Table 1.

At society level, the epidemiology of drug-use-related disease imposes a relevant public-health problem for many countries (Schlenger et al., 1992). For example, almost 40 %

Table 1: Diseases and conditions related to drug use

Anaemia
Bone infections
Central nervous system infections (tetanus, meningitis)
Chronic liver disease
Endocarditis
Heart disease (hypertension, heart attack, stroke)
Hepatitis
Human Immunodeficiency Virus (HIV)
Mental disorders (schizophrenia)
Other retroviruses (HTLV)
Pneumococcal bacteremia
Respiratory diseases
Sexually transmitted diseases (gonorrhea, syphilis)
Skin infections
Tuberculosis

Sources: adapted from Haverkos and Lange (1990), Umbricht-Schneiter et al. (1994), Kaku and Lowenstein (1990)

of reported AIDS cases in the European Union are related to the intravenous use of drugs (European Centre for the Epidemiological Monitoring of AIDS, 1994).

This chapter focuses on the health-care costs of drug-use-related diseases. The treatment and care of these – mostly infectious – diseases require the input of various health-care resources such as hospital beds, nursing care, specialist consultations, day care and GP visits, corresponding to financial budgets and expenditures. Since 1990, research into the spread of HIV/AIDS in groups of intravenous drug users has increased the awareness of these type of costs (Rice et al., 1991). For example, an overview of hospital costing studies of HIV/AIDS was published in 1993 (Postma et al., 1993). In addition, quality standards for these type of costing studies were defined (Tolley and Gyldmark, 1993). As a next step, these studies have been used to assess and project national totals for the use of health-care resource and costs for HIV/AIDS (Rovira and Leidl, 1995; Postma et al., 1995; Kyriopoulos et al., 1992). Furthermore, in scenario analysis, the epidemiological and economic impact of different future courses of the epidemic and different treatment and prevention options are explored (Postma et al., 1993; Rovira and Leidl, 1995). Obviously, in the field of quantitative AIDS research and scenario analysis in particular, a set of new methods has been developed that might be applicable to problems of health-care costs of other drug-use-related infectious diseases (see Chapter 11). This chapter investigates this applicability, and reviews some previous and ongoing research into the health-care costing of drug-use-related diseases.

Chapter 11 also explains the basic framework for scenario analysis. The steps taken are: to construct a conceptual model; to go from a conceptual model to a mathematical model; to integrate into the model epidemiological and health-care information from a structured review; and to construct a reference and several alternative scenarios (Van Genugten et al., 1996). Whereas for non-infectious diseases assessment of health-care burdens can sometimes do without an actual epidemiological model, the nature of infectious diseases necessitates the use of a sometimes complex mathematical model for the spread and progression of the disease (see Chapter 10). The methodological parts of this chapter focus on approaches for linking epidemiological models with health-care information, preferably specified for drug users. To enable specific estimates for drug users, this chapter emphasises the need for patient-based data on health-care use. Several detailed examples are presented here, including one that addresses some of the problems that arise when different costing studies for one disease are reviewed and compared, focusing on hepatitis costs. A second example illustrates the use of patient-based data for estimating health-care impacts for drug users. The final example specifies the drug-use-related hospital care in the reference scenario for AIDS hospital care in EU countries (Postma et al., 1997).

Linking epidemiological models to health-care information

The objective of modelling the health-care impact of drug use is to assess current and project future health-care needs and financial requirements. The general design of the model links an epidemiological sub-model to formatted information on health-care needs. The epidemiological sub-model can be any type that assesses and projects the

burden of disease at national or multinational (for example, EU) level. A common distinction in infectious-disease modelling is between population-dynamic models and empirical models (Bailey and Heisterkamp, 1995). Empirical models draw on surveillance data or counted new cases of a disease. Population-dynamic models explicitly incorporate some of the transmission dynamics of the infectious agent. To enable detailed country-specific impact assessments, the generalisability of the epidemiological model to various country contexts has become an important prerequisite. In particular, the model should enable the provision of standardised epidemiological information of comparable quality for all EU countries.

The format of the information on health-care needs should be in line with the type of output of the epidemiological model. The format of health-care-needs information will be addressed in the next section. Generally, the impact calculus is based on incidence or prevalence output figures. The methods are respectively labelled as incidence-based and prevalence-based (Hartunian and Smart, 1981). The first concerns accounting lifetime health-care resources to the year of incidence of disease. The second method concerns an annualisation of these lifetime health-care resources. To cover the full economic impact of a health-care measure, it is necessary to analyse the lifetime consequences for resource needs incurred by a representative case. This incidence-based approach does, however, assume that future health-care technologies will remain the same as in the analysis, and also extends all other assumptions up to the time until the last incident case has been dealt with. If the assumption of everything staying equal holds true for the period of analysis, and if the full economic impact is what is being referred to in a decision-making context (for example, in decisions on research budgets), the incidence-based approach is the more appropriate one. If, on the other hand, the budget impact of a health-care technology is the focus of analysis, and if this impact is expected to change with new treatment technologies, then the prevalence-based approach, applied over a well-defined period of time, will be the more appropriate approach. Obviously, the incidence- and prevalence-based approaches may differ when applied to the same epidemiological situation. Ideally, a modelling approach can deal with both incidence- and prevalence-based assessments.

Model development

Besides the availability of an appropriate epidemiological sub-model, a full model for the health-care impact of drug use requires a sound procedure for incorporating reliable estimates of health-care needs and costs. This procedure should take into account a disease-staging that distinguishes stages that differ significantly with respect to resource needs. Neglecting to take account of such stages can introduce serious biases in future projections, in particular in situations with non-stable epidemiology and changing treatment patterns (Postma et al., 1998). This staging may or may not overlap with clinically relevant stages. For example, for chronic diseases a useful staging has proved to be one that distinguishes a late or terminal phase from one or several other stages (Scitovsky, 1984). In particular, this concept has been of value in modelling the costs of a disease as clinically diverse and complex as HIV/AIDS. For a clinically clearer-cut disease progression, the use of a progression tree is a good alternative. Antoñanzas et

al. (1995) present a progression tree for the clinical evolution of hepatitis. Hepatitis infection has an acute symptomatic manifestation and may result in chronic carriership. The symptomatic form is often mild, but can be severe or fulminant in some cases with serious associated liver disease possibly requiring transplantation. The chronic carrier might develop active hepatitis, associated with cirrhosis, heptocellular carcinoma and again possibly requiring liver transplantation. In the next step, reliable parameter estimates for resource use and financial costs for the distinguished stages in the progression tree have to be specified.

In health economics, several classifications are used for expressing the consequences of disease on resource use and costs: direct versus indirect; patient-related versus general programme; and health-care versus non-health-care costs (Rovira, 1990). Direct costs arise from the treatment and care of patients in different health-care settings and from research, public-information, prevention and media campaigns. Indirect costs relate to production losses as a result of premature death (potential years of life lost or PYLL) or illness (sickness absence). Patient-related costs are directly associated with the individual person at a certain disease stage (Johnson, 1988). These costs comprise direct costs of treatment and care as well as indirect costs. General programme costs are the costs of research, prevention, public information and the maintenance of services for the benefit of the entire population. In particular, patient-related costs are very suitable for linking to epidemiology in an incidence- or prevalence-based assessment.

Drug use involves a range of clinical and public-health problems with alleged cost implications. Costing implications will vary from country to country and from health-care sector to health-care sector. Drug-use-related disease, such as HIV/AIDS and hepatitis, involves hospital care, home care, medication, and so on. Furthermore, non-patient-related costs arise from providing information programmes for drug users, providing appropriate protection from infection for health-care workers, screening pregnant women for HIV and hepatitis, and screening and heat treating blood and blood products. Finally, mortality due to AIDS and hepatitis complications gives rise to life years lost and corresponding production losses. In this chapter, the main focus is on the patient-related direct costs of drug-use-related disease in the hospital sector.

As both HIV/AIDS and hepatitis involve serious complications and complex health-care interventions that take place several years after the actual infection with the virus, the concept of discounting future costs is important. Discounting reflects a weighting procedure, where costs are weighted less the farther in the future they accrue. The discounting procedure reflects time preference: most people would rather get 100 Euros now than, for example, next year. First, 100 Euros acquired now can be invested and hence grow to an amount that is more than 100 Euros next year. Second, there is uncertainty with regard to future happenings, namely, will it be possible to benefit from the 100 Euros next year as now? There is widespread consensus among health economists on the discounting of costs (Drummond et al., 1987; Gold et al., 1996) and the rationale for the procedure can be found in the existence of 'time preference', opportunity cost of investment and economic growth. There is less agreement on the choice of the appropriate value of the discount rate, although it should clearly be related to the marginal rates of return on investment and market interest. Discount rates varying from

zero to 20 % have been reported (Petrou et al., 1993). Furthermore, discounting should not be confused with correcting costs for inflation rates. In addition to discounting, costing studies should apply a deflator to express costs in different years in a currency of constant price level. It should be noted that there is less consensus on discounting health benefits, such as PYLL and quality-adjusted life years (Mooney and Olsen, 1994; Parsonage and Neuberger, 1992; Keeler and Scretin, 1983). A full discussion of the controversy is beyond the scope of this chpater, but the general recommendation is to include in the model calculation options with and without discounting future health benefits.

The reliability of the parameter estimates of the direct health-care costs depends on the quality and representativeness of the data that are used for estimation. Specific criteria for the data can refer to the number of patients in the database/costing study (also in relevant sub-groups, for example drug users), the number of health-care centres involved, their distribution throughout the country and a set up of the data-gathering allowing in-depth analysis for fine-tuning information to the epidemiological model's design. To enhance comparability of HIV/AIDS costing, a standardised approach has been developed (Tolley and Gyldmark, 1993; Tolley et al., 1998), identifying five specific fields for standardisation.

- First, information should be detectable from a sufficiently large patient sample size in the relevant disease stages and sufficient patient follow-up for recording resource use and costs to be ensured.
- To produce representative estimates for national and regional planning purposes a sufficient number of the main treatment sites covering the country/region needs to be included.
- It should be possible to break down aggregates so as to identify the contribution of different components in the care and cost totals.
- The possibility should be built in to produce homogeneous cost groups, for example using several classifications for HIV/AIDS as defined by the Centres for Disease Control in Atlanta, Georgia, the international classification of diseases (ICD) codes, the Turner stages (Turner et al., 1991) or specific disease-staging definitions, such as terminal/late versus chronic.
- Finally, a standardised methodology for approach and presentation should be followed to enable the effect of 'true factors' to be identified (differences in epidemics, health-care structure and practice, pricing structures and local health-care labour markets). The standardised approach is intended to provide guidelines in these five areas and could possibly serve as a clearing house for studies of resource use and costs for HIV/AIDS and other diseases, such as hepatitis.

Model applications

Reviewing costing studies

The standardised approach for costing studies is in essence a hierarchy of technical 'accounting' steps associated with estimating mean resource use and costs per person-year for disease stages, specific episodes and lifetime resources and costs. The accounting framework is based on a number of elements of standard economic costing practice, and can be applied to any disease. Five main standards were used to develop the approach.

- As a starting point, the perspective of the evaluation should be clarified, in other words, the hospital or the health-services planner at the national or multi-national (e.g. EU) level.
- Next, cost components and resource-use units should be identified according to the perspective adopted. The framework covers four standard cost components: inpatient; out-patient; day care (inpatient for several days with no overnight stay); and day case (admitted and discharged in one day). For specific purposes the last two or three components are taken together as day care and out-patient care respectively.
- To measure physical resource use, rigorous methods of study design and data collection are essential, such as a predominantly patient-based (bottom-up) assessment of resources.
- Resources should be valued using appropriate methods of unit-cost (or price) estimation that best reflect opportunity cost. Therefore, caution must be exercised in using health-service charges, tariffs or prices.
- Mean resource and cost estimates should be presented according to the decision-making context for which they are to be applied (total, disease-stage, clinical condition or transmission group such as drug users). Together, the accounting framework and the standard principles on which it is based, if adopted in new cost studies, should produce cost estimates that meet the normative criteria of good quality and comparability better than do previous studies.

The health-care impact of hepatitis has primarily been modelled for the purpose of evaluating the costs and benefits of vaccination programmes. Since both plasma-derived and recombinant-DNA-produced vaccines have become available, many studies in this field have been performed. In this framework several costing studies for hepatitis were conducted in the 1990s (Jefferson and Demicheli, 1994; Badia et al., 1997). In general, concern has been expressed about the quality and comparability of such disease-specific health-care costing studies (Drummond et al., 1993). This concern is reinforced by Jefferson and Demicheli (1994) who indicate a wide range for direct health care (25th percentile US$ 650, 75th percentile US$ 3 768) and indirect costs (25th percentile US$ 14 164, 75th percentile US$ 127 592) in different hepatitis-costing studies. The high level of the indirect costs can be explained by the use of the human-capital method for valuing years of life lost to hepatitis. This section provides some of the costing details

of selected cost-effectiveness analyses of hepatitis B vaccination in EU countries published in international health-economics journals (Table 2; Antoñanzas et al., 1995; Beutels et al., 1996; Kerleau et al., 1995). In this study, cost figures were converted into US$ using the exchange rates for the local currencies as reported in the publications. Obviously, for better comparison, health-care-specific purchasing-power parities could be considered (OECD, 1993). Costs were not deflated, but cost estimates refer to different price years. Already this limited selection of three studies indicates major differences in the way the various cost categories are conceived. For example, the Spanish study (Antoñanzas et al., 1995) calculates significant costs for the stage of chronic infection, in other words, healthy carriership versus active carriership. The costs for assessing active chronic infection include those for a physician's visit, liver biopsy, interferon treatment, eight-day hospitalisation, six follow-up visits and a final follow-up visit with testing for hepatitis B markers. The other two studies neglect or assign lower cost figures to these stages. Furthermore, the Belgian (Bentels et al., 1996) and French (Kerleau et al., 1996) studies apply much higher cost estimates for the follow-up and long-term complications of active chronicity. The higher estimate for Belgium compared to France can partly be explained by the higher discount rate in the latter study.

Table 2: Reference, country of study, price level, discount rate and costs in US$ for various hepatitis stages

Reference	Antoñanzas et al. (1995)	Beutels et al. (1996)	Kerleau et al. (1996)
Country	Spain	Belgium	France
Price level	1990	1995	1990
Discount rate	5 %	5 %	8 %
	Costs per stage		
i[a]	560	420	1 150
ii	3 660	2 990	5 400
iii	1 510	3 560[b]	–
iv	6 310	4 870[b]	–
v	115 000[c]	83 000	80 000
vi	510	310	–
vii	8 080	1 230	–
viii	210	130	150
ix	420	1 430	5 000
x	1 470	11 100	4 600
xi	–	15 300	10 000

Notes: i-vii are expressed per stage; viii-xi are expressed per full year of follow-up

[a] i: acute, non-fulminant mild; ii: acute, non-fulminant severe; iii: acute, fulminant fatal; iv: acute fulminant fatal; v: liver transplant; vi: assessment of healthy chronicity; vii: assessment of active chronicity; viii: healthy carrier; ix: active carrier; x: cirrhosis; xi: cancer

[b] own estimates

[c] including the present value of future follow-up of liver-transplant patients

– indicates not reported

A standardised approach to cost estimation would help explain such differences in cost estimates, maximising the usefulness of the estimates for a range of policy purposes at various decision-making levels. Obviously, there is a need for a standardised approach to estimate the per-patient, per-stage and total costs of hepatitis treatment and care. The development of this approach should follow principles outlined in earlier work on costing standardisation of HIV/AIDS hospital care (Tolley and Gyldmark, 1993; Tolley et al., 1998). Kerleau et al. (1995) pays special attention to hepatitis B cost-effectiveness for interventions among drug users. In particular, the authors estimate benefit-to-costs ratios of 3.5 and 5.7 for respectively vaccinating all drug users and vaccinating only those drug users found susceptible to hepatitis (after screening).

Selecting cost studies for scenario analysis

As mentioned above, the standardised approach to health-care costing has been applied to HIV/AIDS. For the purpose of linking epidemiology and health care in future scenarios, specific databases on AIDS hospital care in the EU have been selected. Three criteria were applied. First, information should be recent (1990 or later), reflecting current efficiency of practices in hospital care. During the 1980s and into the 1990s, efficiency changed due to increasing experience with managing people with AIDS (PWAs). Furthermore, resource use has been controlled by more (cost-)effective use; for example, the substitution of less expensive out-patient care for hospital inpatient care. Second, selected studies should include data that can be assumed to be representative of the national situations. The representativeness was considered safeguarded in multi-centre studies. Third, the resource-use data should be patient-based, enabling resource use to be linked to individual patients, patient groups and disease stages, such as HIV infection and specific opportunistic illnesses. Six studies – most of them designed as databases – satisfied most of these criteria (individual studies are listed in Postma et al., 1997), and are theoretically able to specify separate estimates for drug-use-related HIV/AIDS. In a previous application no distinction between different risk groups was made (Postma et al., 1997). For this chaper, it was possible to specify estimates for drug users from Dutch (four hospitals) and Italian databases (ten hospitals).

It is well known that hospital-resource needs incurred by AIDS vary in the different stages of the disease. In particular, the terminal phase of AIDS entails a relatively high intensity of resource needs and costs. Therefore, two stages are differentiated for PWAs: a final stage of six months (maximum) before death (late stage: LS); and a stage for the foregoing period (chronic stage: CS). It has been argued that neglecting this concept threatens the validity and comparability of estimates derived from different studies of use and introduces significant biases in consequent future projections (Postma et al., 1998). Applying the staging concept to Greece, Spain, France, Italy, the Netherlands and the UK has shown that per-person-year (ppy) inpatient days in LS are at least twice those in CS. For example, in Italy and the Netherlands ppy inpatient days in LS are, respectively, 60 and 80, whereas the respective CS figures are 24 and 33 inpatient days.

Table 3 shows the Italian and Dutch figures, specified according to drug-use-related and non-drug-use-related AIDS. Italy and the Netherlands show opposite patterns. In Italy, total hospital contacts are higher for drug-related AIDS in both stages. In particular, late-stage inpatient-day needs are relatively high. Obviously, the late stage of drug-related AIDS in Italy involves a relatively frequent occurrence of inpatient-day-intensive complications, such as AIDS dementia. In the Netherlands, the figures illustrate a different pattern; LS inpatient-day needs are relatively low. This is consistent with a previous Dutch study (Postma et al., 1995) showing relatively low use of ppy inpatient days for drug users (16.1 versus 47.9 for homosexual men). However, in this study no correction for chronic and late stage could be made. Furthermore, both the current and the previous Dutch studies involve small sample sizes for drug users.

AIDS scenarios

In a mixed incidence/prevalence-based approach, epidemiological information (Postma et al., 1997) is linked with the per-person-year hospital-resource needs of PWAs. Estimated ppy hospital-resource use is used as a proxy for resource needs, implicitly assuming that recorded resource use reflects actual care needs. Annual resource needs for CS are estimated by multiplying the annual period-prevalence in CS by the appropriate ppy resource-need parameters for CS. Annual resource needs for LS are estimated by multiplying annual mortality, the duration of stay in LS (six months or less) and the appropriate ppy resource-need parameters. It is assumed that ppy resource

Table 3: Hospital resource use per person-year of AIDS, AIDS stage and risk group, Italy and the Netherlands

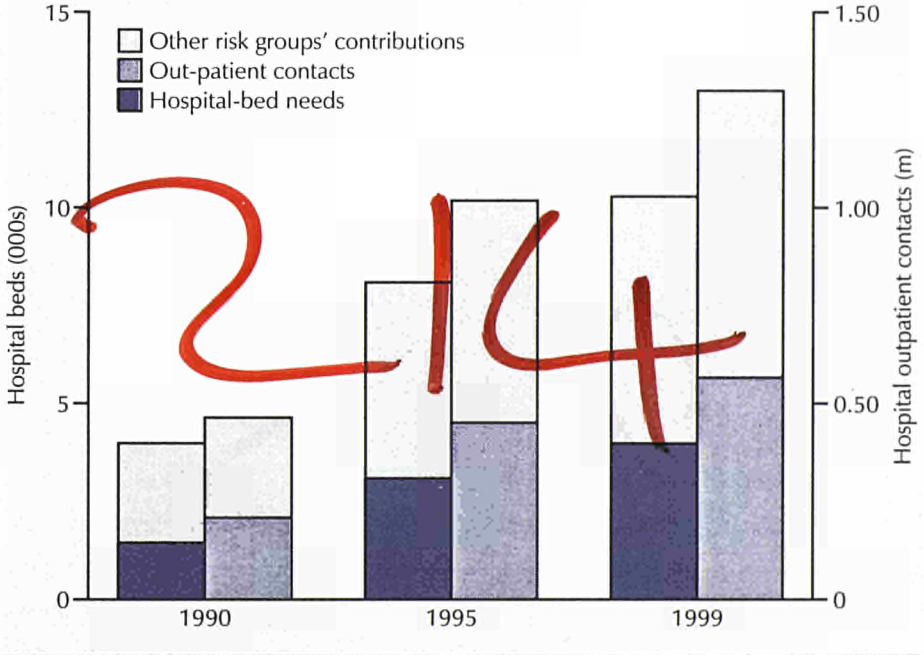
Country	Risk group	Stage	IPD[a]	OPC	THC
Italy	all	CS[b]	33	44	77
		LS	60	28	88
	drug-use-related	CS	32	46	78
		LS	68	30	98
	non-drug-use-related	CS	34	39	73
		LS	40	26	66
Netherlands	all	CS	24	21	45
		LS	80	23	103
	drug-use-related	CS[c]	27	19	46
		LS[c]	30	33	63
	non-drug-use-related	CS	23	21	44
		LS	82	22	104

Notes: [a] IPD: inpatient days; OPC: out-patient contacts; THC: total hospital contacts
[b] CS: chronic stage; LS: late stage
[c] This figure was derived from only three patients
Source: Postma et al. (1997)

needs remain stable over calendar time in each stage so that the effect of changes in epidemiology can be isolated. Finally, a generalisation had to be chosen to extrapolate the approach from the six countries with high-quality hospital-resource-use studies to those lacking these studies. This generalisation specifies the calculation of average ppy resource needs in CS and LS for the countries with the information, and these are applied to the countries with missing information.

Figure 1 presents period prevalence-based estimates of needs for hospital beds and out-patient hospital contacts (including day-care treatment) in the reference scenario. The crucial epidemiological assumption in the reference scenario specifies that the future level of annual HIV incidence equals the level estimated for 1993. Other scenarios are presented elsewhere (Postma et al., 1997). Figure 1 indicates that less than 40 % of EU hospital-bed needs concern drug users, whereas they are responsible for over 40 % of the need for out-patient contacts. This gives rise to the hypothesis that the share of out-patient treatment in total hospital care is relatively high for drug users. Overall, however, the share of drug-user-related hospital care is below what might be expected on the basis of their share in incidence. This share rises to more than 50 % in 1999. The difference between the shares in hospital care and incidence can be explained by the fact that drug-use-related AIDS is primarily

Figure 1: Reference scenario for drug-use-related hospital-resource needs

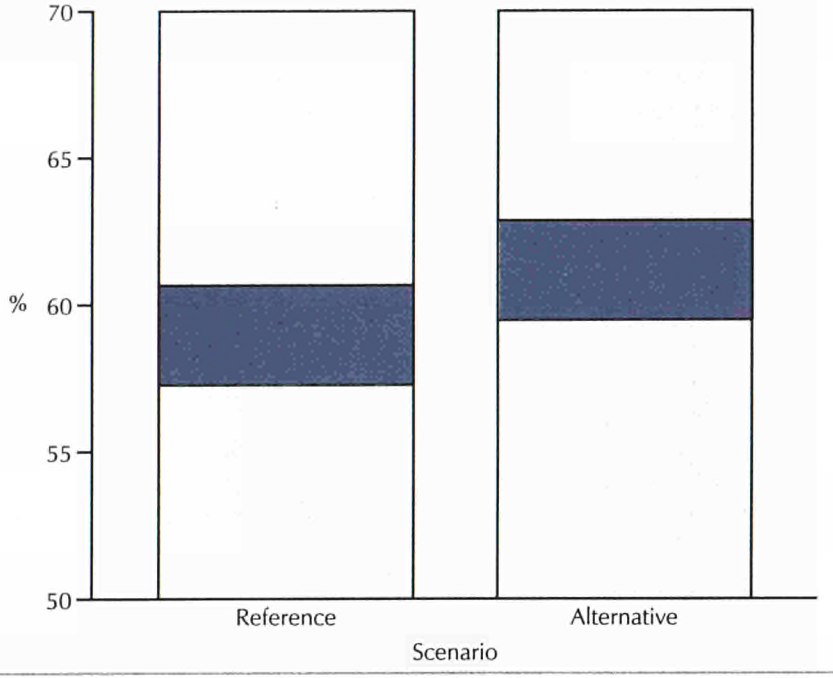


Note: Hospital-bed needs and outpatient contacts indicated as parts of totals for all risk groups
Source for totals: Postma et al. (1997)

located in the southern European countries (for example, almost 70 % of the new AIDS cases in Spain in 1999 were drug users, compared to just over 60 % in 1995). These countries show on average lower estimates of ppy inpatient days and out-patient contact needs than the EU averages of 54 days and 22 contacts. For example, in Italy estimated ppy inpatient-day needs are 38.5 for 1995 and the need for out-patient contacts in Spain amounts to 15.6 ppy of AIDS in 1995.

As Italy is one of the EU countries with a large share of drug-use-related AIDS in the total epidemic, Figure 2 specifies the 1995 needs for total hospital contacts (inpatient days, day-care days and out-patient contacts) of drug users in this country (incidence-based calculations) (Postma et al., 2000). Furthermore, total hospital contacts are shown taking the relatively high resource needs of drug-related AIDS cases into account (Table 3). Finally, the total estimate for all risk groups is shown (including homosexual men and heterosexuals). The figure indicates that approximately 60 % of all hospital contacts in 1995 are estimated to have been for drug users.

Figure 2: Total hospital contacts (inpatient days, out-patient contacts and day-care days), Italy, 1995



Notes: [a]percentage related to drug use in total for all risk groups. The dark blue band indicates the range between lower and upper 95% confidence limits
[b]reference: see Figure 1; alternative: taking into account relatively high resource needs for drug-related AIDS cases

Source: Postma et al. (1997, 2000)

Critical assessment of models

Obviously, this type of costing analysis requires a rather detailed patient-based registration that is ideally set up to provide information for cost-of-illness and future projections. Furthermore, the type of information required can only be retrieved from bottom-up, patient-based data on health-care resource use and costs in cases of time-related staging of disease with documentation over the full life span of a specific patient. This detailed information enables estimates to be provided of drug-use-related disease in the present and in future projections to support health-policy decisions. They provide a quantitative order of magnitude for health-care resources in this area. They specify the patient mix, for example in terms of drug user versus non-drug user and chronic versus late, which are important indicators for planning hospital staff, capacity and support.

A standardised approach to costing health care has helped in designing bottom-up, patient-based costing studies for HIV/AIDS that can render the appropriate information. If such an approach is followed, the information that is necessary for cost assessment and projection is identifiable. Further application of the standardised approach could involve the examination of local, national and multinational variations in cost estimates for drug-use-related diseases such as HIV/AIDS and hepatitis. If bias due to methodological differences is minimised, cost variations due to differences in clinical practice, health-care structure and organisation and epidemiology can be more readily identified. In addition, the standards should improve investigation of resource-use variations over time in one or more sites. Furthermore, detailed analysis in specific areas of treatment and care could facilitate investigation of the cost-effectiveness of alternative policy and care options in these areas (for example, screening for HIV and hepatitis, HIV combination triple therapy and hepatitis vaccination). Finally, the standardised approach could provide a starting point for further development to cover non-hospital care and social care.

The focus of the models and model applications discussed above is mainly on the direct health-care costs. One study attempts to estimate the indirect costs of drug use (French et al., 1996). Using a quality-adjusted life-year methodology (French et al., 1992), the value of avoiding drug-use-related morbidity and mortality is estimated. As a common unit of well-being applicable to all illnesses and diseases, quality-adjusted life years are calculated using the Rosser/Kind index (Rosser and Kind, 1978). The Rosser/Kind index uses eight disability levels and four distress levels to define 29 health states. Perfect health corresponds to a value of 1, death corresponds to a value of 0. For example, a moderately acute case of hepatitis B corresponds to a value of 0.956 (severe social and slight work disabilities and moderate distress). These life years are valued by means of the 'US dollar value' for human life. In this particular case, a 'dollar value' of human life of US $5 million is applied (Viscusi, 1993). Different from the human capital method that only values the present value of future earnings, the value of pain and suffering is included in the calculations. For example, the dollar values for AIDS, severe hypertension and ten weeks of severe pneumonia are estimated at respectively US $115 000, US $75 000 and US $58 000 (Viscusi, 1993). The total lifetime value for HIV/AIDS amounts to US $158 000 (including pre-AIDS stages). Furthermore, including mortality costs, the estimated full dollar value of avoiding a single case of HIV equals

US $2 655 000. For comparison, lifetime health-care costs for HIV/AIDS in the United States have been estimated at US $119 000 (Hellinger, 1993), which obviously only reflects a small part of all the costs of a drug-related HIV/AIDS case.

Costing impact by scenario analysis as outlined above represents an exercise that should potentially underpin health-care policy for drug users. However, due to differences in patient group, type of hospital care provided, and stage of disease dealt with, the quantification of current and future costing impact is quite a complex task. By integrating recent and detailed data, important steps can be made in improving the current state of information. Nonetheless, uncertainty will remain, leaving issues of availability of representative data on drug-use-related health-care costs, impact on non-hospital care sectors and dynamic integration of treatment technology for future work. In the short term, however, an indication of the order of magnitude of the impact of drug-use-related disease on EU health-care systems can be provided, probably indicating that drug use should be an important issue on the agendas of health-care policy-makers in the coming years.

Conclusions

Essentially, this chapter argues that future costing scenarios for drug-use-related health-care costs can be built by applying sound costing principles and by standardised epidemiological modelling using relevant disease-staging.

As a first step, standardised gathering of health-care and costing information is crucial. This involves applying classifications of health-care costs, breaking down aggregate cost figures in line with conceptual models of the disease and recalculations for standard discount rates, price levels and currencies. It is therefore recommended that the standardisation of costing data collection as developed for HIV/AIDS is extended to other drug-related diseases, in other words hepatitis as the next step.

For hepatitis, this chapter summarises some of the work that has been undertaken on costing the disease, in particular regarding the framework of cost-effectiveness analyses of vaccination strategies. As a logical start, it is recommended that existing databases for hepatitis care and costing be reviewed, in particular their appropriateness for application in drug-use-related costing scenarios should be evaluated. For example, the French AIDS-care database (Postma et al., 1997) includes hepatitis care, and could potentially be used for French drug-use-related hepatitis costs.

Future scenario analysis of health-care and costing impacts requires the application of epidemiological models – primarily of infectious diseases – in combination with health-economic information as indicated above. An appropriate choice for disease staging is important to avert structural biases in health-care needs assessments and projections. This has been argued for HIV/AIDS above, and has recently been illustrated for chronic diseases as well (Postma et al., 1998). In combination with major differences in resource needs and associated costs between these stages, epidemiological dynamics have major implications for resource needs and costs. In consequence, any application of costing approaches and scenario analysis should control for this disease staging. Again, it is

recommended to extend methods developed for HIV/AIDS – in particular the disease-staging approach – to other drug-use-related disease costing (Postma, 1998).

Furthermore, this chapter illustrates that detailed information on drug-use-related health-care needs and costs for HIV/AIDS is potentially available by citing some detailed information for drug-use-related AIDS in Italy and the Netherlands. It is recommended that for other countries the acquisition of this type of information be pursued. For example, Spanish, English and French AIDS-care databases (Postma et al., 1997) allow the extraction of this information. This information will enable the estimation of drug-use-related health-care costs in the EU at present and in future scenarios. As a result it is recommended that ultimately future costing scenarios are developed for all relevant drug-use-related diseases.

Acknowledgement

The authors would like to thank A. M. Tramarin and M. G. W. Dijkgraaf for kindly providing the databases on drug-use-related hospital care for AIDS in, respectively, Italy and the Netherlands.

References

Antoñanzas, F., Garuz, R., Rovira, J., Antón, F., Trinxet, C., Navas, E., Salleras, L. (1995). 'Cost-effectiveness analysis of hepatitis B vaccination strategies in Catalonia, Spain'. *Pharmacoeconomics* 7: 428–43.

Badia, X., Nocea, G., Rovira, J. (1997). 'Differences in the methodology and data of economic evaluations of a health programme; the case of hepatitis B vaccination programmes in Spain'. *Pharmacoeconomics* 2: 175–83.

Bailey, N. T. J., Heisterkamp, S. H. (eds) (1995). *Preliminary review on the mathematical models and data used for the baseline analysis*. Bilthoven: National Institute of Public Health and the Environment (RIVM).

Beutels, P., Tormans, G., van Damme, P., van Doorslaer, E. (1996). 'Algemene vaccinatie tegen hepatitis B in Vlaanderen: kosten-effectief?' *Tijdschrift voor Sociale Gezondheidszorg* 74: 272–81.

Dinwiddie, S. H., Reich, T., Cloninger, C. R. (1992). 'Psychiatric comorbidity and suicidality among intravenous drug users'. *Journal of Clinical Psychiatry* 53: 364.

Drummond, M. F., Brandt, A., Luce, B., Rovira, J. (1993). 'Standardising economic evaluation methodologies in health care: practice, problems and potential'. *International Journal of Technology Assessment in Health Care* 1: 26–36.

Drummond, M. F., Stoddart, G. L., Torrance, G. W. (1987). *Methods for economic evaluation of health care programmes*. Oxford: Oxford University Press.

European Centre for the Epidemiological Monitoring of AIDS (1994). *AIDS surveillance in the European Community and COST countries*. Quarterly Report 29. Paris: European Centre for the Epidemiological Monitoring of AIDS.

French, M. T., Mauskopf, J. A. (1992). 'A quality-of-life method for estimating the value of avoided morbidity'. *American Journal of Public Health* 82: 1553.

French, M. T., Mauskopf, J. A., Teague, J. L., Roland, E. J. (1996). 'Estimating the dollar value of health outcomes from drug-abuse interventions'. *Medical Care* 34 (9): 890–910.

Gold, M. R., Siegel, J. E., Russel, L. B., Weinstein, M. C. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Hartunian, N. S., Smart, C. N. (1981). *The incidence and economic costs of major health impairments*. Washington DC: National Academy Press.

Haverkos, H. W., Lange, R. W. (1990). 'Serious infections other than human immunodeficiency virus among intravenous drug abusers'. *Journal of Infectious Diseases* 161: 894.

Hellinger, F. (1993). 'The lifetime cost of treating a person with HIV'. *Journal of the American Medical Association* 270: 474.

Jefferson, T., Demicheli, V. (1994). 'Is vaccination against hepatitis B efficient? A review of world literature'. *Health Economics* 3: 25–37.

Johnson, A. M. (1988). 'Economic aspects of care and prevention of HIV infection and AIDS', in Jager, J. C., Ruitenberg, E. J. (eds) *Statistical analysis and mathematical modelling of AIDS*. Oxford: Oxford University Press.

Kaku, D. A., Lowenstein, D. H. (1990). 'Emergence of recreational drug abuse as a major risk factor for stroke in young adults'. *Annals of Internal Medicine* 113: 821.

Keeler, E. B., Scretin, S. (1983). 'Discounting of life-saving and other non-monetary effects'. *Management Science* 29: 300–6.

Kerleau, M., Flori, Y. A., Nalpas, B., Lanoé, J. L., Berthelot, P., Fardeau-Gautier, M. (1995). 'Analyse coût-avantage d'une politique de prévention vaccinale de l'hépatite virale B'. *Revue d'Epidémiologie et de Santé Publique* 43: 48–60.

Kyriopoulos, J., Georgoussi, E., Gennimata, D., Niakas, D., Skoutelis, G. (1992). 'AIDS cost in Greece; 1989–1993 – macroeconomic approach and estimations'. *Health Review* 4: 35–8.

Mooney, G., Olson, J. A. (1994). 'QALYs: where next?' in McGuire, A., Fenn, P., Meyhew, K. (eds) *Providing health care*. Oxford: Oxford University Press, 120–40.

Organisation for Economic Co-operation and Development (OECD) (1993). *OECD health systems*. Paris: OECD.

Parsonage, M., Neuberger, H. (1992). 'Discounting and health benefits'. *Health Economics* 1: 71–6.

Petrou, S., Malek, M., Davey, P. G. (1993). 'The reliability of cost-utility estimates in cost-per-QALY league tables'. *Pharmacoeconomics* 3: 345–53.

Postma, M. J. (1998). *Assessment of the economic impact of AIDS at national and multi-national level*. Zutphen: Koninklyke Wöhrmann bu.

Postma, M. J., Kornarou, H., Paparizos, V., Leidl, R. M., Tolley, K., Kyriopoulos, J., Jager, J. C. (2000). 'Hospital care for persons with AIDS in European Union countries; a cross-country comparison'. *Health Care Management Science* 3: 1–7.

Postma, M. J., Jager, J. C., Ruwaard, D., van Loy, N. C. M., Leidl, R. M. (1998). 'Disease-staging for modelling current and future health-care impact of disease: illustrations for diabetes and AIDS'. *Health Policy* 43: 45–54.

Postma, M. J., Tolley, K., Leidl, R. M., Downs, A. M., Beck, E. J., Tramarin, A. M., Flori, Y. A., Santin, M., Antoñanzas, F., Kornarou, H., Paparizos, V. C. C., Dijkgraaf, M. G. W., Borleffs, J., Luijben, A. J. P., Jager, J. C. for the European research team on AIDS scenarios (1997). 'Hospital care for persons with AIDS in the European Union'. *Health Policy* 41: 157–76.

Postma, M. J., Jager, J. C., Dijkgraaf, M. G. W., Borleffs, J. C. C., Tolley, K., Leidl, R. M. (1995). 'AIDS scenarios for the Netherlands: the economic impact on hospitals'. *Health Policy* 31: 127–50.

Postma, M. J., Leidl, R. M., Downs, A. M., Rovira, J., Tolley, K., Gyldmark, M., Jager, J. C. (1993). 'Economic impact of the AIDS epidemic in the European Community: towards multinational scenarios on hospital care and costs'. *AIDS* 7 (5): 41–53.

Rice, D. P., Kelman, S., Miller, L. S. (1991). 'Costs of alcohol and drug abuse and mental illness, 1985–88'. *Public Health Reports* 106: 280–92.

Rosser, R., Kind, P. (1978). 'A scale of valuations for status of illness: is there a social consensus?' *International Journal of Epidemiology* 7: 347.

Rovira, J. (1990). 'Economic aspects of AIDS', in Schwefel, D., Leidl, R., Rovira, J., Drummond, M. F. (eds) *Economic aspects of AIDS and HIV infection.* Berlin: Springer-Verlag.

Rovira, J., Leidl, R. (1995). 'Projecting individual healthcare costs of HIV/AIDS patients in Catalonia', in Fitzsimons, D., Hardy, V., Tolley, K. (eds) *The economic and social impact of AIDS in Europe.* London: Cassell.

Schlenger, W. E., Kroutil, L. A., Roland, E. J., Dennis, M. L. (1992). *National evaluation of models for linking drug abuse treatment and primary care; descriptive report of phase one findings.* Rockville MD: National Institute on Drug Abuse.

Scitovsky, A. A. (1984). '"The high costs of dying": what do the data show?' *Milbank Quarterly* 62: 591–608.

Tolley, K., Gyldmark, M. (1993). 'The treatment and care costs of people with HIV infection or AIDS: development of a standardised cost framework for Europe'. *Health Policy* 24: 55–70.

Tolley, K., Postma, M. J., Jager, J. C., Good, M. A. (1998). 'Standardised costing framework for HIV/AIDS hospital care in the European Union', in Kyriopoulos, J., et al. (eds) *AIDS cost analysis and patient classification.* Athens: Exandax, 37–51.

Turner, B. J., Markson, L. E., McKee, L., Houchens, R., Fanning, T. (1991). 'The AIDS-defining diagnosis and subsequent complications: a survival-based severity index'. *Journal of Acquired Immune Deficiency Syndromes* 4: 1059–71.

Umbricht-Schneiter, A., Ginn, D. H., Pabst, K. M., Bigelow, G. E. (1994). 'Providing medical care to methadone clinic patients: referral vs on-site care'. *American Journal of Public Health* 84: 207.

Van Genugten, M. L. L., Rutten, F. F. H., Jager, J. C. (1996). *Scenario development and costing in health care: methodological accomplishments and practical guidelines.* Utrecht: International Books.

Viscusi, W. K. (1993). 'The value of risks to life and health'. *Journal of Economic Literature* 32: 1912.

CHAPTER 13

ECONOMIC MARKET MODELS

Matthew Sutton

Economists study behaviour in terms of markets or trading situations. In general, that study is divided between the analysis of demand or consumer behaviour and that of supply or producer and distributor behaviour. A fundamental distinction in the economics literature is the level of analysis. There are broadly three levels of analysis:

- picoeconomics: the study of behaviour within individuals (often across time);
- microeconomics: the study of behaviour across individuals within areas (or markets);
- macroeconomics: the study of behaviour across areas (or markets).

An interest in the dynamics of consumption suggests that picoeconomic analysis would be particularly pertinent in the case of illicit drugs, and much of the economic literature on addictive substances concerns picoeconomic relationships. However, whilst theoretical developments in the economics of addiction (and demand for addictive substances) have been more extensive at the picoeconomic level, it is macroeconomic outcomes which command the most policy attention. Much of the existing data on market outcomes and prices/purities are at the aggregate market level (e.g. Pompidou Group, 1994). The lack of conceptual guidance at this level tends to make interpretation of trends in these data precarious and the evaluation of policy interventions rather depressing (Caulkins and Reuter, 1996; Weatherburn and Lind, 1997).

One of the central premises of economics is that market prices will change to ensure that the supply of, and demand for, commodities reaches equilibrium. Therefore, much applied work in economics focuses on the relationship between quantities traded and market prices. The other fundamental lesson from economics is that choices are made within a budget constraint and therefore income levels will affect quantities purchased in the market.

Economic-market models can be used to generate estimates of the determinants of dynamic patterns of drug prevalence/incidence. There are very few applications of these models either within or outside Europe. The demand-side aspects of the models are based on theoretical models of individual behaviour which easily incorporate contextual/environmental factors. However, preferences play a major role in the determination of demand and there has been relatively little economic work on the formation of preferences.

The supply side of drug markets is known to vary substantially between areas, even within countries. It would therefore be difficult to extrapolate results to other areas (particularly between the United States and Europe) because of the differences in the supply sides of markets. A study of how drug markets are formed and how their structure develops would aid the generalisability of any economic-market applications.

Economic-market models require good-quality data on levels of consumption or expenditure, prices, purities and total income. The analysis can either be at individual or market level. At individual level, there are considerable problems in the collection of consumption data, and measures of market prices and purities may have little explanatory power for individual outcomes because of wide variability between individuals in prices paid and purities received. Moreover, recent developments in dynamic modelling of individual consumption require panel data on quantities consumed and/or prices paid. Despite the obvious difficulty with the feasibility of such data collection, recent attempts to collect economic data at individual level are encouraging.

At a market level, the vast majority of economic data are derived from enforcement-agency reports. Price data often rely on reports from undercover buys and purities relate to drugs seized. Neither data source has been established to be representative. Moreover, the problems with inferences about levels of consumption based on 'market indicators' such as seizures, drug-related deaths or registrations for treatment are well known. Based on such data, applied studies have found little link between economic factors, enforcement activity and market outcomes.

This chapter contains an overview of the existing theoretical and applied economic analysis of drug markets. Available studies on demand are considered, followed by a limited analysis of the supply side of the market, a review of addiction models and finally some comments on market equilibrium. Some conclusions are drawn on the feasibility of further research across Europe.

Analysis of the demand for drugs

The extent to which demands respond to changes in price and income are termed the price elasticity and income elasticity of demand, respectively. The majority of applied-economics work on addictive substances therefore attempts to estimate price and income elasticities (Leung and Phelps, 1993; Godfrey, 1994). The study of illicit drugs has focused to a greater degree on the price elasticity of demand, and particularly on the way in which addiction may influence the nature of the relationship between prices and demand (Wagstaff and Maynard, 1988; Becker et al., 1992).

However, there is more to economic models of demand and supply than price analysis and it is important to remember that the traditional study of market equilibrium (the downward-sloping demand curve and upward-sloping supply curve) is based on the *ceteris paribus* assumption (everything else held constant). Before reviewing the literature on price effects, a summary of economic work on the other determinants of demand and supply will be presented. Throughout the review, reference is made to applications of new models to the analysis of tobacco and alcohol consumption. Whilst simple

extrapolation of these results to the consumption of illicit drugs is not appropriate, these studies demonstrate the potential for economic methods. Fuller reviews of the economic literature on addictive goods can be found in Godfrey (1994), Leung and Phelps (1993) and Buck et al. (1996).

Non-price determinants of demand

Various drug-using behaviours have been analysed in demand studies. For the consumption of legal addictive substances, applied-economic analyses have focused on the separation of decisions about whether to consume and, given the choice to consume, how much to consume (Atkinson et al., 1984; Jones, 1989). At the macro level, this can be analysed as the percentage of individuals participating in the market and the total level of consumption (Jones, 1989). More recently, Jones (1995) has focused on the determinants of the decision to quit smoking. He finds that individuals with higher potential benefits from quitting (i.e. heavier consumers) are more likely to attempt to quit.

Analyses of the consumption of health-damaging commodities should be set within the context of individuals' decisions about their level of health (Birch and Stoddart, 1990). Grossman's (1972) model of individual health production dominated the early economic literature on individuals' choices about health and lifestyles. Grossman conceptualised health as being a stock variable which depreciates through time and can be replenished with health investments. Inputs into the health-production process include medical care and lifestyle variables, such as exercise and consumption of health-affecting commodities, for example, food and addictive substances. Grossman suggested that individuals with a higher level of education would have better knowledge of the production process and would therefore make more appropriate decisions about health inputs. Education may also affect individuals' knowledge of the health risks of addictive goods. Viscusi's (1992) empirical results suggest that, if anything, smokers overestimate the risks posed to their health by smoking. Viscusi's results have been challenged by Schoenbaum (1997), who found that heavy smokers, in particular, underestimated the risks they were taking with their future health.

Differences in levels of demand between individuals which cannot be explained by these factors are often attributed to differences in preferences. Despite the importance of such considerations, there has been relatively little focus within economics on the ways in which such preferences are formed (Pollak, 1976). Within the study of health-damaging and addictive goods there are some notable exceptions. There has been interest in the effect of habit formation on intertemporal consumption patterns (Alessie and Kapteyn, 1991) and the development of rational addiction models can be seen as a refinement of this broad approach (Becker and Murphy, 1988). Moreover, in the addiction area, there have been developments in the ways in which time preference is formed and manipulated (Becker and Mulligan, 1993). Finally, the effects of peer behaviour on consumption patterns have also received prominence in the area of addiction (Norton et al., 1997). Each of these three developments will be discussed in turn.

Habit formation

Becker and Murphy (1988), for example, propose a theoretical model of addiction which emphasises the time-dependence of consumption levels. Previous theories had incorporated the effect of previous consumption on current consumption, but Becker and Murphy's theory predicted that individuals would be rational, in the sense that they would consider the consequences of current consumption decisions on future time periods. Becker and Murphy's model of rational addiction has been applied to US cocaine-consumption data by Grossman and Chaloupka (1998). Despite the limitations of the data set and the way in which the consumption-level data were collected, Grossman and Chaloupka did find some evidence to support Becker and Murphy's predictions.

Time preference

One of the other central areas of interest in economics has been the way in which individuals discount future occurrences (Buck et al., 1996). Many theoretical models of dynamic behaviour have been based on stable models of intertemporal behaviour (such as constant exponential discounting). Differences in time preference between individuals was one of the main explanations given by Becker and Murphy (1988) for why some individuals would begin an addiction career whilst others would not. Bretteville-Jensen (1997), however, presents empirical estimates of the time preferences of current users, ex-users and never-users. Her results do not suggest that ever-users (i.e. current and ex-users) have systematically different (higher) time-discount rates. An alternative explanation for why current users do display higher time-preference rates as expected, but that ex-users have similar rates to never-users, is that of endogenous time-preference (Becker and Mulligan, 1993). If consumption of certain commodities (such as addictive goods) influences an individual's valuation of the future, drug users may be expected to have shorter planning horizons.

Peer effects

Most economic applications have included peer behaviour as an exogenous determinant of individual behaviour (Lewit et al., 1981; Jones, 1995; Sutton and Godfrey, 1995). Manski (1993) distinguishes between three different causal mechanisms which would give rise to similar patterns of behaviour within groups. Correlated effects arise when individuals form groups on the basis of common characteristics or attitudes. Exogenous effects relate to environmental or group factors which have a common effect on members of the same group. Endogenous effects, on the other hand, pertain to situations in which one individual's behaviour has a direct causal influence on the behaviours of others within the same group. It is only this latter effect which gives rise to the social multiplier which lies behind much of the policy concern with peer effects on behaviour.

Having distinguished between these three types of effect, Manski (1993) demonstrates the difficulty of separating out these effects in cross-sectional data. Rice and Sutton (1997) categorise various economic theories of why individuals' preferences and

behaviours tend to conform in terms of whether they describe endogenous, exogenous or correlated effects. They apply Manski's (1993) framework to the analysis of alcohol-consumption patterns within households and find significant evidence of correlated (or group-selection) effects. They also find evidence of endogenous and/or exogenous effects within households, but are unable to distinguish between them. In a US setting, Norton et al. (1997) find no evidence of correlated effects in groups defined by school attendance.

Analysis of the supply side of the market

Applied analyses of the supply side of illicit drug markets have been almost exclusively linked to the effectiveness of enforcement interventions in the market. Grizzle (1979) did find some evidence of the effectiveness of enforcement. Weatherburn and Lind (1997) attempted to link enforcement activity with market outcomes. They correlated levels of enforcement seizures with price data in the Sydney heroin market. Allowing for time lags between seizures and market outcomes, they found no evidence of an effect of enforcement on market conditions. Apart from being rather depressing for enforcement agencies, it is rather unclear what can be done with these findings. It could be, of course, that the results are a consequence of the data-collection procedure (i.e. enforcement is effective but its effectiveness was not measured) or the level of aggregation at which the analysis was performed (i.e. enforcement is effective but its effectiveness is unmeasurable at this level). More importantly, however, if the results are to be trusted, the study gives no guidance as to *why* enforcement is not effective at the margin. This would require a more detailed understanding of the operation of the supply side of the market.

Wagstaff and Maynard (1988) illustrated the potential effectiveness of enforcement activity in the UK heroin market using Lewis et al.'s (1985) description of the structure of the London heroin market and Polich et al.'s (1984) model of the way in which enforcement activity impacts on the market. There were no real estimates available of the important parameters in the model and therefore the study became a way to generate data demands rather than an applied study of the effectiveness of enforcement. Subsequent studies in the US and the UK (Cave and Reuter, 1988; Crawford et al., 1988; Rydell and Everingham, 1994; Sutton and Maynard, 1994) suffer from similar requirements to base effectiveness estimates on a number of parameter assumptions.

Other studies of the supply side of the market have been broadly descriptive. Dorn et al. (1992), for example, provide a taxonomy of supply organisations which is impossible to operationalise in an economic model of the market. Chatterton et al. (1995) set out to characterise existing descriptions of trading organisations along a number of dimensions. An attempt was made to develop measures for some of these dimensions and to pilot data collection. Based on self-completion questionnaires from individuals attending treatment, an idea of the relative importance of different geographical market areas was sought. Respondents were also asked about the nature of trading relationships between themselves and the next level up in the supply chain. The vast majority of respondents reported using the same dealer either always or usually. Such a pilot exercise

indicates that at least at the lowest level, collection of quantitative data on market structure is feasible, although no tests of validity, reliability or representativeness were attempted.

Addiction and the effect of price on demand

Wagstaff and Maynard (1988) reviewed the pre-rational-addiction literature on models of the effect of addiction on the price elasticity of demand for addictive goods. In the late 1960s and early 1970s, it was widely believed that demand curves for addictive goods were perfectly inelastic, in other words that the quantity demanded by an individual did not respond in any way to changes in price.

However, later additions to the debate proposed a range of options that were available to drug users in the face of changes (particularly increases) in price: switching to close substitutes; entry into treatment; quitting drug use altogether; or reducing consumption through decreasing frequency of use and tolerating withdrawal effects (Holahan, 1973; Bernard, 1983). Blair and Vogel (1983) suggested that demand may be responsive to price at low prices, but that once addicted individuals had reduced their consumption to maintenance doses, further price increases would not decrease consumption any more. Alternatively, White and Luksetich (1983) proposed that demand would become price-responsive at high prices because of difficulties with raising sufficient funds, changes in the relative attractiveness of treatment and increased risks of detection by enforcement agents. Wagstaff and Maynard (1988) suggested that these arguments related to different segments of the demand curve and suggested that this curve may be kinked, with elastic segments at the top and bottom and a perfectly inelastic segment in the middle price range.

However, it was not clear whether this hypothesis referred to individuals or the market as a whole. The inelastic segment referred to individuals on maintenance doses and there was no discussion of whether this range would be common across individuals. Recent empirical work by Bretteville-Jensen and Sutton (1996b) failed to find evidence of changes in price-responsiveness along the market-demand curve. However their results, in common with those of van Ours (1995), suggested that the demand for heroin may be considerably more price-elastic than previously thought. The debate about the price-responsiveness of different groups of users is mirrored in the economic literature on alcohol consumption (Manning et al., 1995; Sutton and Godfrey, 1995).

The level of analysis remains particularly important in analysing the economic determinants of demand patterns. Several studies have demonstrated the wide variability in prices paid by individuals within the same market (Bretteville-Jensen and Sutton, 1996b; Caulkins and Reuter, 1996; Weatherburn and Lind, 1997). Collection of data from individual drug users has been shown to be feasible (Johnson et al., 1985; Grapendaal, 1992; Chatterton et al., 1995; Bretteville-Jensen and Sutton, 1996b) and offers far greater scope than simple analyses of unreliable indicators of market quantities and prices.

Demand, supply and market equilibrium

Many studies of the price responsiveness of the consumption of addictive commodities refer to the price elasticity of demand rather than consumption. However, economic models emphasise that demand remains unobserved in most applications and all that is recorded is the quantity traded between demanders and suppliers. In markets where the price mechanism operates as assumed (i.e. it brings about equilibrium between supply and demand), the relationship between price and quantity traded can be assumed to relate to demand if supply is either held constant or controlled for using supply variables in the analysis. Where prices do not move to give equilibrium, more complex analysis is required such as disequilibrium models (Maddala, 1983; Chapter 10, above). Only structural relationships between market variables can be assumed to be generalisable between areas, and therefore careful consideration must be given to the relationship between demand, supply, prices and market equilibrium. The nature of these problems will differ considerably depending on whether the analysis is at the market or individual level.

Analysis of the supply and demand sides of the illicit drug market is further complicated by the fact that many users are also small-scale dealers. Bretteville-Jensen and Sutton (1996a) found significant differences between dealers and non-dealers in their levels of consumption and the ways in which they generated income. In further analysis, they found that the ways in which market factors affected the behaviour of these two groups also differed markedly. At this level, they found significant effects of prices, income levels and potential profit opportunities on individual behaviour. The structural relationship between supply and demand decisions by individuals has also been discussed by Sommers et al. (1996).

Conclusions

Paucity of data is one of the principal reasons why economic-market models have not been extensively applied in Europe. However, there are other contributory factors. Compared to other disciplines, economists have had relatively little involvement in the analysis of drug use and related problems. Economists' involvement in illicit-drug debates has been primarily to contribute semantically to the libertarian debate about drug legalisation. It is probably fair also to suggest that the credibility of economic market models suffers from both over-simplification by many commentators and the 'non-results' found in the applied analyses undertaken at market level.

Given the lack of applications of these models, it is tempting to propose that *any* research project in this area would have considerable 'added value'. However, to be worthwhile, an economic analysis of illicit drug markets would require considerable investment and a particular blend of researchers. The economic input should be firmly grounded in theory, but its thrust should be the application of existing concepts. The 'added value' of market analysis would primarily be a better understanding of interactions between policy interventions and market outcomes, since existing policy evaluations either ignore the role of markets in dictating outcomes (particularly in the context of treatment) or

adopt extremely simplistic notions of the market (in the field of enforcement). Economic analysis is required to identify *structural* relationships between different aspects of the market.

There is no restriction on economic-market models as to the size of geographic area at which the analysis can be performed. However, individual-level analysis is the most clearly formulated and aggregate-level analysis may suffer from aggregation problems, such as the ecological fallacy or inappropriate aggregation of non-linear relationships at individual level. There is economic literature on the definitions of market areas and market segmentation which could be used to define the pertinent geographic area for analysis.

A successful research project in this area would require a substantial amount of resources to ensure sufficient data of good quality on levels of consumption, prices and purities. This would require economic input into the design of the data collection, but also fieldwork by interviewers sympathetic to economic methods and concepts. The subsequent analysis and interpretation of the data would need an economist conscious of the special (illicit) nature of the phenomenon under study and the consequently inevitable trade-off between the importance of academic rigour and the need to obtain policy-relevant results. At this time, it would be unlikely that an economist with sufficient specialist knowledge could be found and the project would therefore need to be multidisciplinary.

References

Alessie, R., Kapteyn, A. (1991). 'Habit formation, interdependent preferences and demographic effects in the almost ideal demand system'. *Economic Journal* 101: 404–19.

Atkinson, A., Gomulka, J., Stern, N. (1984). *Household expenditure on tobacco 1970–1980: evidence from the Family Expenditure Survey*. Economic and Social Science Research Council Programme, Taxation, Incentives and the Distribution of Income, No. 57. London: London School of Economics.

Becker, G. S., Mulligan, C. B. (1993). 'On the endogenous determination of time preference'. Unpublished manuscript.

Becker, G. S., Murphy, K. M. (1988). 'A theory of rational addiction'. *Journal of Political Economy* 96 (4): 675–700.

Becker, G. S., Grossman, M., Murphy, K. M. (1992). 'Rational addiction and the effect of price on consumption', in Loewenstein, G., Elster, J. (eds) *Choice over time*. New York: Russell Sage Foundation 361–70.

Bernard, G. (1983). 'An economic analysis of the illicit drug market'. *International Journal of the Addictions* 18 (5): 681–700.

Birch, S., Stoddart, G. (1990). *Incentives to be healthy: an economic model of health-related behaviour*. Discussion Paper 63. York: Centre for Health Economics.

Blair, R. D., Vogel, R. J. (1983). 'Heroin addiction and urban crime'. *Public Finance Quarterly* 1 (4): 457–67.

Bretteville-Jensen, A. L. (1997). 'Addiction and discounting'. Paper presented to the 18th Nordic Health Economists' Study Group, Helsinki.

Bretteville-Jensen, A. L., Sutton, M. (1996a). 'The income-generating behaviour of injecting drug-users in Oslo'. *Addiction* 91 (1): 63–79.

Bretteville-Jensen, A. L., Sutton, M. (1996b). *Under the influence of the market: an applied study of illicitly selling and consuming heroin*. Discussion Paper 143. York: Centre for Health Economics.

Buck, D., Godfrey, C., Sutton, M. (1996). 'Economic and other views of addiction: implications for the evaluation of policy options'. *Drug and Alcohol Review* 15: 357–68.

Caulkins, J., Reuter, P. (1996). 'The meaning and utility of drug prices'. *Addiction* 91 (9): 1261–4.

Cave, J. A. K., Reuter, P. (1988). *The interdictor's lot: a dynamic model of the market for drug smuggling services*. Santa Monica, CA: RAND.

Chatterton, M., Gibson, G., Gilman, M., Godfrey, C., Sutton, M., Wright, A. (1995). *Performance indicators for local anti-drug strategies: a preliminary analysis*. Crime Detection and Prevention Series Paper 62. London: Home Office Police Research Group.

Crawford, G. B., Reuter, P., Isaacson, K., Murphy, P. (1988). *Simulation of adaptive response: a model of drug interdiction*. Santa Monica, CA: RAND.

Dorn, N., Murji, K., South, N. (1992). *Traffickers: drug markets and law enforcement*. London: Routledge.

Godfrey, C. (1994). 'Economic influences on changes in population and personal substance behaviour', in Edwards, G., Lader, M. (eds) *Addiction: processes of change*. Society for the Study of Addiction Monograph No. 3. Oxford: Oxford University Press.

Grapendaal, M. (1992). 'Cutting their coat according to their cloth: economic behaviour of Amsterdam opiate users'. *International Journal of the Addictions* 27 (4): 487–501.

Grizzle, G. A. (1979). 'Illicit drug availability – effects of a local law-enforcement program'. *Evaluation Quarterly* 3 (4): 516–36.

Grossman, M. (1972). 'On the concept of health capital and the demand for health'. *Journal of Political Economy* 80: 223–55.

Grossman, M., Chaloupka, F. J. (1998). 'The demand for cocaine by young adults: a rational addiction approach'. *Journal of Health Economics* 17 (4): 427–74.

Holahan, J. (1973). 'The economics of the control of the illegal supply of heroin'. *Public Finance Quarterly* 1 (4): 467–77.

Johnson, B. D., et al. (1985). *Taking care of business: the economics of crime by heroin abusers*. Lexington, MD: Lexington Books.

Jones, A. M. (1995). *A microeconometric analysis of smoking in the UK Health and Lifestyle Survey*. Discussion Paper 139. York: Centre for Health Economics.

Jones, A. M. (1989). 'A double-hurdle model of cigarette consumption'. *Journal of Applied Econometrics* 4: 23–39.

Leung, S. I., Phelps, C. E. (1993). '"My kingdom for a drink...?": a review of estimates of the price sensitivity of demand for alcoholic beverages', in Hilton, M. E., Bloss, G. (eds) *Economics and the prevention of alcohol-related problems*. Washington DC: US Department of Health and Human Services, 1–31.

Lewis, R., Hartnoll, R., Bryer, S., Daviaud, E., Mitcheson, M. (1985). 'Scoring smack: the illicit heroin market in London'. *British Journal of Addiction* 80: 281–90.

Lewit, E. M., Coate, D., Grossman, M. (1981). 'The effects of government regulation on teenage smoking'. *Journal of Law and Economics* 24: 545–70.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.

Manning, W. G., Blumberg, L., Moulton, L. H. (1995). 'The demand for alcohol: the differential response to price'. *Journal of Health Economics* 14: 123–48.

Manski, C. F. (1993). 'Identification of endogenous social effects: the reflection problem'. *Review of Economic Studies* 60: 531–42.

Norton, E., Lindrooth, R., Ennett, S. (1997). 'Controlling for endogeneity of peer substance use on adolescent alcohol and tobacco use'. Paper presented to the Sixth European Workshop on Econometrics and Health Economics, New University of Lisbon, Lisbon.

Polich, J. M., Ellikson, P. L., Reuter, P., Kalion, J. P. (1984). *Strategies for controlling adolescent drug use*. Santa Monica, CA: RAND.

Pollak, R. A. (1976). 'Interdependent preferences'. *American Economic Review* 66: 309–20.

Pompidou Group (Co-operation Group to Combat Drug Abuse and Illicit Trafficking in Drugs) (1994). *Multi-city study: drug misuse trends in thirteen European cities*. Strasbourg: Council of Europe Press.

Rice, N., Sutton, M. (1997). *The identification, estimation and interpretation of 'higher-level effects': the problem of endogeneity*. Technical Paper 7. York: Centre for Health Economics.

Rydell, C. P., Everingham, S. S. (1994). *Controlling cocaine: supply versus demand programs*. Santa Monica, CA: RAND.

Schoenbaum, M. (1997). 'Do smokers understand the mortality effects of smoking? Evidence from the Health and Retirement Survey'. *American Journal of Public Health* 87 (5): 755–9.

Sommers, I., Baskin, D., Fagan, J. (1996). 'The structural relationship between drug use, drug dealing, and other income support activities among women drug sellers'. *Journal of Drug Issues* 26 (4): 975–1006.

Sutton, M., Godfrey, C. (1995). 'A grouped data regression approach to estimating economic and social influences on individual drinking behaviour'. *Health Economics* 4 (3): 237–47.

Sutton, M., Maynard, A. (1994). *Trends in the cost-effectiveness of enforcement activity in the illicit heroin market, 1979–91*. YARTIC Occasional Paper 4. York: Centre for Health Economics.

van Ours, J. C. (1995). 'The price-elasticity of hard drugs – the case of opium in the Dutch East Indies, 1923–1938'. *Journal of Political Economy* 103 (2): 261–79.

Viscusi, W. K. (1992). *Smoking: taking the risky decision*. Oxford: Oxford University Press.

Wagstaff, A., Maynard, A. (1988). *Economic aspects of the illicit drug market and drug enforcement policies in the United Kingdom*. Home Office Research Study 95. London: Home Office.

Weatherburn, D., Lind, B. (1997). 'Drug law enforcement policy and its impact on the heroin market'. *Addiction* 92: 557–69.

White, M. D., Luksetich, W. A. (1983). 'Heroin: price elasticity and enforcement strategies'. *Economic Inquiry* 21: 557–64.

General conclusions

GENERAL CONCLUSIONS

Richard Hartnoll and Lucas Wiessing

Modelling has played and continues to play a principal role in various research fields such as the economy and public health, especially where macro-level decision-making is concerned. Scenario analyses of future trends in demography, AIDS cases or pension costs, for example, help to foresee future needs for services and financial reserves. Cost–benefit and cost-effectiveness analyses of alternative interventions help to decide whether procedures such as hepatitis vaccination or breast-cancer screening are best carried out for the general population or best limited to high-risk groups.

In the drugs field, only a limited amount of such work has been carried out, even though the potential for some basic approaches had already been demonstrated in the 1970s and earlier. It seems almost paradoxical that techniques that are meant to make sense of scarce data have rarely been applied in Europe because of a lack of data. An additional reason may be that European drugs research and social sciences in general have put stronger emphasis on historical or qualitative approaches than, for example, in the United States. Also, the European field of drugs research itself was very limited in size and resources until quite recently and is still under-funded. However, in recent years, the field has been developing rapidly and opening up to other disciplines including mathematics, statistics and economics. Data availability and quality are improving and the potential to use models to forecast future trends or to estimate hidden processes, such as initiation of drug use, has consequently been increasing.

Quantitative techniques such as modelling should not be applied in isolation. A multidisciplinary approach is essential, especially for modellers, who are often highly expert in mathematics or statistics but may sometimes have less 'field knowledge' of drug use. As the projects that led to this monograph, and their ongoing follow-ups have shown, the most useful results of modelling are derived by inter-disciplinary teams of mathematicians, statisticians, epidemiologists, economists, sociologists and the medical doctors who see addicted patients every day.

At the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), the most obvious complementary element to the European modelling network, of which this monograph is an important output, is the European network of qualitative researchers (see http://www.qed.org.uk and *Understanding and responding to drug use: the role of qualitative research*. EMCDDA Scientific Monograph No. 4. Lisbon: European Monitoring Centre for Drugs and Drug Addiction, 2000). Like the modelling network, this qualitative network aims to promote tools and methods to improve understanding of processes, hidden populations and usefulness of responses, be it using totally different and highly complementary approaches to mathematical and statistical modelling. It is very important, both for qualitative and quantitative drug-use researchers in Europe, that closer collaboration and mutual understanding of each other's approaches develops.

Future directions

The EMCDDA has already embarked on a cluster of modelling projects as a follow-up to the initial reviewing project that resulted in this monograph. These activities are centred on three broad aims:

- bringing together and consolidating European research infrastructure by forming networks of researchers, funding meetings for exchange of results and stimulating collaborations and joint publications;
- developing priority themes on quantitative epidemiological research of drug use by funding small pilot or 'seed' studies meant to be continued by national and more robust EU research funding;
- linking the results of the networks and pilot projects to the needs of policy-makers by holding joint seminars, focusing research objectives around policy relevant questions and presenting results in formats that are understandable by non-specialists.

Bringing together and consolidating European research infrastructure

This includes stimulating research initiatives such as writing proposals for larger research funding, holding meetings where experts can exchange results and keeping the research community informed of new initiatives through presentations at scientific meetings and conferences and publication in the EMCDDA *Annual report on the state of the drugs problem in the European Union*, its newsletter *Drugnet Europe* and its website, http://www.emcdda.org. The EMCDDA-funded European modelling network, that started with the reviewing project which resulted in this monograph, was consolidated at the end of 1998 through a research grant from the EU Commission under the fourth framework programme of former Directorate-General XII, Targeted socio-economic research (TSER). The EMCDDA took the initiative for writing the research proposal and is coordinating the project. This is a so-called 'network project' where funding is mainly for regular meetings to exchange results on ongoing work. Six working groups are active on the following topics: national prevalence estimation; local prevalence estimation; geographic spread; time trends and incidence; costs and cost-effectiveness; and drug markets and policy options. The EMCDDA/TSER project has resulted in a series of meetings on different themes with participation from over 40 experts in modelling and other disciplines. Outcomes to date include more than 30 scientific publications that have directly benefited from the project (see http://www.emcdda.org/activities/epidem_tser.shtml).

Developing priority themes on quantitative epidemiological research

From the 1997 review meetings in York and the May 1998 Lisbon seminar, a number of themes emerged that appeared the most relevant to policy-making in the short term as well as feasible given the data and expertise available. Given the lack of an overall macro picture of drug use in Europe in 1998, the main emphasis was placed on descriptive epidemiology. Already in 1996 a seminar held in Strasbourg had highlighted the importance of improving methods for estimating the prevalence of problem drug use, using separate

approaches for the local and the national levels (see *Estimating the prevalence of problem drug use in Europe*. EMCDDA Scientific Monograph No. 1. Lisbon: European Monitoring Centre for Drugs and Drug Addiction, 1997). Additional themes that emerged from the modelling meetings in York and Lisbon were: estimation of the incidence of problem drug use using back-calculation methods and drug-treatment data; linking geographic information systems to forecasting models to investigate and map the geographic spread of problem drug use; using economic methods to calculate the health-care and social costs of drug-related infections and the cost-effectiveness of interventions; modelling drug markets and supply-side factors to gauge the potential for policy options aimed, for example, at increasing the price of drugs. As descriptive epidemiology improves and more data become available, new themes can concentrate further on understanding mechanisms and interrelationships at a more detailed level and in a more fundamental way. At the initial meetings, modelling micro-diffusion processes was not recommended as it was seen as too difficult to find appropriate data. However, more basic analyses are already proving important, as a prerequisite for understanding developments at macro level.

Linking the results of the networks and pilot projects to the needs of policy-makers

Even if there is an obvious need for a centre such as the EMCDDA, it has been challenging to make the link between the interests and priorities of quantitative researchers and policy-makers. This was related to the themes themselves, the nature of the methods used, the types of participants in different meetings and their interests and 'languages', and also to the difficulty of clearly defining the needs of policy-makers and the concept of 'policy-maker' itself. While the importance of prevalence estimation of problem drug use seemed obvious from the beginning and never raised many questions as to its relevance, the wider concept of 'modelling' has not always been fully understood by non-modellers in the drugs field. Most of the methods that underlie the work of estimating prevalence, incidence or cost-effectiveness are hard to understand for non-specialists while the experts themselves may often not be used to writing for an audience other than expert readers. Mixing specialists and non-specialists at larger meetings proved important but not easy, as the different jargons used in some cases resulted in Babylonian discussions. On the other hand, in both York meetings, which were held mostly among modellers, discussions often converged on the working of different techniques rather than on the highest priorities in drugs policy, reflecting the methodological interests of the experts involved.

In conclusion, it is clear that work remains to be done in explaining the importance of using quantitative methods to estimate prevalence, incidence, spread, costs and cost-effectiveness and the interplay of supply factors in drug markets to non-specialist decision-makers. It may sometimes be difficult to understand that to obtain a sound evidence base in the longer term, some patience is necessary and funding is needed to develop methodological infrastructure. It bodes well for the future, however, that despite these difficulties a strong network of European modellers has been formed. Results are improving quickly in relevance and quality, in parallel to the increasing access to data sources among networks of other experts. It is also encouraging that European drug-strategy documents are increasingly based on setting quantitative and measurable goals. It is only a matter of time before the drugs field will be using these techniques as if they had always been there.

Contributors

CONTRIBUTORS

Peter W. Achterberg
National Institute of Public Health and the Environment (RIVM)
Department of Public Health Forecasting
PO Box 1 Bilthoven
3720 BA Bilthoven
The Netherlands
Tel. (31-30) 274 38 37
Fax (31-30) 274 44 50
E-mail: peter.achterberg@rivm.nl

Azzedine Boumhgar
Observatoires Regionaux de la Santé
Téléperformance France
12 rue Charles Fourier
F-75013 Paris
Tel. (33) 155 76 40 10
Fax (33) 155 76 40 20
E-mail: a.boumhgar@teleperformance.fr

Catherine Comiskey
Department of Mathematics
National University of Ireland
Maynooth
Co. Kildare
Ireland
Tel. (353-1) 708 39 94
Fax (353-1) 708 39 13
E-mail: cc@maths.may.ie

Martin Frischer
Department of Medicines Management
University of Keele
Staffordshire ST5 5BG
UK
Tel. (44-1782) 58 35 68
Fax (44-1782) 71 35 86
E-mail: m.frischer@keele.ac.uk

Christine Godfrey
Centre for Health Economics
University of York
Heslington
York YO10 5DD
UK
Tel. (44-1904) 43 45 07
Fax (44-1904) 43 27 00
E-mail: cg2@york.ac.uk

Richard Hartnoll
EMCDDA
Rua da Cruz de Santa Apolónia 23–25
P-1149–045 Lisbon
Tel. (351) 218 11 30 06
Fax (351) 218 13 79 43
E-mail: Richard.Hartnoll@emcdda.org

Gordon Hay
Centre for Drug Misuse Research
12–15 Western Court
Glasgow G12 8SQ
UK
Tel. (44-141) 330 54 13
Fax (44-141) 330 28 20
E-mail: gkua24@udcf.gla.ac.uk

Heath Heatlie
Department of Medicines Management
Keele University
Staffordshire ST5 5BG
UK
Tel. (44-1782) 58 35 62
Fax (44-1782) 71 35 86
E-mail: h.f.heatlie@keele.ac.uk

Matthew Hickman
Centre for Research on Drugs and Health Behaviour
Department of Social Science and Medicine
Imperial College School of Medicine
St Dunstans Road
London W6 8RP
UK
Tel. (44-20) 75 94 07 74
Fax (44-20) 75 94 08 52
E-mail: m.hickman@ic.ac.uk

Johannes Jager
National Institute of Public Health and the Environment (RIVM)
PO Box 1
3720 BA Bilthoven
The Netherlands
Tel. (31-30) 274 25 16
Fax (31-30) 274 44 66
E-mail: Hans.Jager@rivm.nl

James Kahan
RAND Europe
Newtonweg 1
2333 CP Leiden
The Netherlands
Tel. (31-71) 524 51 51
Fax (31-71) 524 51 91
E-mail: kahan@rand.org

Ludwig Kraus
Institut für Therapieforschung (IFT)
Parzivalstr. 25
D-80804 München
Tel. (49-89) 36 08 04 30
Fax (49-89) 36 08 04 49
E-mail: kraus@ift.de

Mirjam Kretzschmar
National Institute of Public Health and the Environment (RIVM)
PO Box 1
3720 BA Bilthoven
The Netherlands
Tel. (31-30) 274 37 49
Fax (31-30) 274 44 09
E-mail: mirjam.kretzschmar@rivm.nl

Fabio Mariani
Department of Epidemiology
Institute of Clinical Physiology
National Research Council
Via Moruzzi
I-56100 Pisa
Tel. (39-050) 315 20 94
Fax (39-050) 315 20 95
E-mail: marifa@ifc.pi.cnr.it

Roberto Mollica
Treatment Outpatient Addiction Unit (SERT)
ASL Città di Milano
District 5
381 via Delle Forze Armate
I-20100 Milan
Tel. (39-02) 48 92 02 71
Fax (39-02) 48 92 63 09
E-mail: mollicar@tin.it

Steve Parrott
Centre for Health Economics
University of York
Heslington
York YO10 5DD
UK
Tel. (44-1904) 43 36 33
Fax (44-1904) 43 36 44
E-mail: sjp22@york.ac.uk

Maarten Postma
Groningen University Institute for Drug Exploration
Groningen Research Institute of Pharmacy (GUIDE/GRIP)
Antonius Deusinglaan 1 9713 AV
Groningen
The Netherlands
Tel. (31-50) 363 26 07
Fax (31-50) 363 27 72
E-mail: m.postma@farm.rug.nl

Carla Rossi
University of Rome Tor Vergata
Dipartimento Di Matematica
Via Ricerca Scientifica
I-00133 Rome
Tel. (39-06) 72 59 46 76
Fax (39-06) 72 59 46 99
E-mail: C.Rossi@agora.stm.it or rossi@axp.mat.uniroma2.it

Gianpaolo Scalia Tomba
University of Rome 'La Sapienza'
Via A. Scarpa 16
I-00161 Rome
Tel. (39-06) 49 91 66 76
Fax (39-06) 495 76 47
E-mail: gianpi@dmmm.uniroma1.it

Roland Simon
Institut für Therapieforschung (IFT)
Parzivalstr. 25
D-80804 München
Tel. (49-89) 36 08 04 40
Fax (49-89) 36 08 04 49
E-mail: simon@ift.de

Matthew Sutton
Department of General Practice
University of Glasgow
4 Lancaster Crescent
Glasgow G12 0RR
UK
Tel. (44-141) 211 16 66
Fax (44-141) 211 16 67
E-mail: M.Sutton@udcf.gla.ac.uk

Colin Taylor
National Addiction Centre
Addiction Sciences Building
4 Windsor Walk
London SE5 8AF
UK
Tel. (44-20) 79 19 34 49
Fax (44-20) 72 52 48 94
E-mail: colin.taylor@iop.bpmf.ac.uk

Keith Tolley
Health Outcomes Research
GlaxoSmithKline UK
Stockley Park
Uxbridge
Middlesex UB11 1BT
UK
Tel. (44-20) 89 90 30 22
Fax (44-20) 89 90 30 15
E-mail: Kt23149@GlaxoWellcome.co.uk

Vanessa Waby
Centre for Health Economics
University of York
Heslington
York YO10 5DD
UK
Tel. (44-1904) 43 26 95
Fax (44-1904) 43 36 44
E-mail: vjw2@york.ac.uk

Lucas Wiessing
EMCDDA
Rua da Cruz de Santa Apolónia 23–25
P-1149–045 Lisbon
Tel. (351) 218 11 30 16
Fax (351) 218 13 79 43
E-mail: lucas.wiessing@emcdda.org

Philip Young
Department of Health Sciences and Clinical Evaluation
University of York
Heslington
York YO10 5DD
UK
Tel. (44-1904) 43 45 11
Fax (44-1904) 43 45 17
E-mail: pjy2@york.ac.uk

Address:
European Monitoring Centre for Drugs and Drug Addiction (EMCDDA)
Rua da Cruz de Santa Apolónia 23–25
P-1149-045 Lisbon

Tel.
(351) 21 811 30 00

Fax
(351) 21 813 17 11

E-mail:
info@emcdda.org

Website:
http://www.emcdda.org

EMCDDA, January 2001

European Monitoring Centre for Drugs and Drug Addiction

EMCDDA Scientific Monograph Series No 6
Modelling drug use: methods to quantify and understand hidden processes

Luxembourg: Office for Official Publications of the European Communities

2001 – 246 pp. – 16 x 24 cm

ISBN 92-9168-056-7

Price (including VAT in Luxembourg): EUR 18

EMCDDA Scientific Monograph Series

15 16 5 AO-18-98-655-EN-C

The European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) was set up in the face of an escalating drug problem in the European Union and a lack of sound and comparable information on the subject at European level. Established by Council Regulation (EEC) No 302/93 on 8 February 1993, the Centre became fully operational in 1995. Its main goal is to provide 'objective, reliable and comparable information at European level concerning drugs and drug addiction and their consequences'.

The Centre's tasks are divided into four categories:

- collecting and analysing existing data;
- improving data-comparison methods;
- disseminating data; and
- co-operating with European and international bodies and organisations, and with non-EU countries.

The EMCDDA works exclusively in the field of information.

Located in Lisbon, the EMCDDA is one of 11 decentralised agencies set up by the European Union to carry out specialised technical or scientific work. As such, the Centre is funded by the Community budget but is autonomous in its operations.

Price (excluding VAT) in Luxembourg: EUR 18

OFFICE FOR OFFICIAL PUBLICATIONS
OF THE EUROPEAN COMMUNITIES

L-2985 Luxembourg

ISBN 92-9168-056-7

9 789291 680566 >