Epidemiol. Infect. (2004), **132**, 1139–1149. © 2004 Cambridge University Press
DOI: 10.1017/S0950268804002997 Printed in the United Kingdom

Definition and estimation of an actual reproduction number describing past infectious disease transmission: application to HIV epidemics among homosexual men in Denmark, Norway and Sweden

E. J. AMUNDSEN[1,2,3]*, H. STIGUM[2], J.-A. RØTTINGEN[4,5] AND O. O. AALEN[3]

[1] *Norwegian Institute for Alcohol and Drug Research* (*SIRUS*), *Oslo, Norway*
[2] *Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway*
[3] *Section of Medical Statistics, University of Oslo, Oslo, Norway*
[4] *Institute for Nutrition Research, University of Oslo, Oslo, Norway*
[5] *Division for Infectious Disease Control, Norwegian Institute of Public Health, Oslo, Norway*

(*Accepted 19 July 2004*)

SUMMARY

Prevalence and incidence measures are the common way to describe epidemics. The reproduction number supplies information on the potential for growth or decline of an epidemic. We define an actual reproduction number for infectious disease transmission that has taken place. An estimator is suggested, based on the number of new infections observed in a given time-interval, the number of those infected at the start of the interval, and the length of the infectious period. That estimator is applied to HIV among men having sex with other men over the period, 1977–1995, in Scandinavia. The actual reproduction number was estimated with acceptable certainty from the period, 1981–1982, yielding a value of 15 secondary cases. A value of less than one secondary case was assessed for the period, 1988–1995, in Denmark and Sweden. The actual reproduction number gives us some additional understanding of the dynamics of epidemics, compared with prevalence and incidence curves.

INTRODUCTION

Common measures for describing epidemics include the prevalence pool, or the number of live cases at a given point in time and the absolute rates, or the number of new cases per time unit [1]. The prevalence measure describes the magnitude of the epidemic, and the absolute rates over time give us information about the development of the epidemic. When both the absolute rates and the prevalence pools are positive, neither measure tells us whether an epidemic is decreasing or increasing. However, measures for increase and decrease of past and recent epidemics can be identified using the concept of reproduction numbers [2–5]. This tool was developed to study the potential spread of an infectious disease.

The basic reproduction number, R_0, is the reproduction number when all subjects are susceptible, i.e. the average number of secondary cases the infection is transmitted to during a typical individual's infectious period in a situation where all persons are susceptible. The effective reproduction number, R_e, is the reproduction number when not all subjects are susceptible. If R_0 is estimated to be >1, the introduction of an infected subject in a population will set off an epidemic. If R_e is >1, an ongoing epidemic will increase.

The ideas behind this concept can be applied to describe an epidemic that has taken place or is developing. Thus, a decrease, for example, has occurred if 'each infected person was replaced by less than one

* Author for correspondence: Dr E. J. Amundsen, Norwegian Institute for Alcohol and Drug Research, PO Box 565 Sentrum, 0105 Oslo, Norway.
(Email: eja@sirus.no)

other infected person during his/her infectious period'. An increase has occurred if 'each infected person was replaced by more than one other infected person during his/her infectious period'.

The distinction between, on one hand, the estimation of the actual spread of an epidemic that has occurred and, on the other hand, the potential for spread in a population where all are susceptible (R_0), or the potential for further spread when some are already infected (R_e), is not always made in the literature. Potterat et al. [6] state that, 'In principle, this basic reproduction number should be estimable from individual level data; that is from observing the number of persons to whom one infected individual spreads the disease.' Potterat applies this concept to *Chlamydia* infection, based on contact-tracing and DNA amplification testing. Haydon et al. [7] also obtained an estimate of the 'case-reproduction ratio' as directly as possible from the data for the 2001 foot-and-mouth outbreak in the United Kingdom. Haydon reconstructed 'epidemic trees' from records on infected properties that might have or did give rise to other infected properties. Borgdorff et al. [8] used DNA 'fingerprints' of all *Mycobacterium tuberculosis* isolates from January 1993 to June 1995 to identify groups of patients that had isolates with identical fingerprints. Borgdorff's group estimated what they called an effective reproduction rate associated with recent transmissions [8].

The literature on such direct estimation of recent increases or decreases of infections is scarce [9]. The main body of the literature on R_e and R_0 relates to the potential for spread and control. Earlier, more simple deterministic mathematical expressions have been replaced by more advanced models, both deterministic and stochastic models [10–18].

In earlier studies, the reproduction numbers for potential development were originally defined as population measures. These were calculated accentuating a person with the 'typical' risk behaviours in the population, for example, in a situation with a heterogeneous mixing of partners. Later models, with distributions of risk behaviours and other parameters, have now been applied, and, thus, the reproduction numbers can now be viewed as means in distributions for individuals.

Estimation of actual reproduction numbers via individual (or subject) reproduction of cases is difficult to achieve since individual disease transmission is not recorded or recordable. Population, or group, measures need to be developed. In the study of HIV epidemics and the actual spread of HIV, it has been practically impossible to establish actual chains of who infected whom. The problem of estimating prevalence pools and absolute rates has been a challenge, in and of itself.

The aim of this article is to define the actual reproduction number for an epidemic that has occurred, to suggest an estimator for this number, and to apply it to HIV among men infected through sex with men in Denmark, Norway and Sweden. The actual reproduction numbers are compared to estimates of absolute rates and prevalence pools from 1977 to 1995, obtained by back-calculation methods. The information and understanding obtained are summarized.

METHODS

We define, fairly loosely, the actual reproduction number, R_a, as the average number of secondary cases per case to which the infection was actually transmitted during the infectious period in a population. A stochastic perspective is applied, and, therefore, we will distinguish between R_a (an underlying true mean) and different estimators for R_a. The ideal design for estimating R_a would be a cohort of subjects infected at the same time. These subjects could be followed over their infectious period, and the number of secondary cases counted. As in earlier published works, group figures can be totalled from individual records. This type of ideal design hardly ever exists, in practice, for infectious diseases, and, therefore, other types of estimators have to be constructed for R_a.

A possible and practical way to establish an estimator follows: In a short time-interval, an estimator of the transmission rate of the infectious agent will, intuitively, be the number of new cases in the interval (the absolute rate) divided by the number of those infected by the start of the interval (the prevalence pool). This estimator would tell one how many new subjects an already infected subject transmitted the infectious agent to, on average, during the given interval. Multiplying this quantity by the average length of the infectious period gives an estimator of how many new cases one subject transmits the infectious agent to, on average, if the situation in the short time-interval is extended to the entire infectious period. A formal expression of this estimator is:

$$\hat{R}_a(i) = (X_i / P_{i-1}) * D, \qquad (*)$$

where X_i is the absolute rate in interval i, P_{i-1} is the prevalence pool at the end of the preceding interval,

and D is the length of the average infectious period. The estimator can only be applied when P_{i-1} is greater than zero. There is no simple expression for its variance even when both X_i and P_{i-1} are normally distributed.

The proposed estimator for R_a in (*) will be a mix of the values for the actual transmission of infections (from the P_{i-1} subjects to X_i new subjects) and potential development in the period, D. The length of the interval, i, should actually be chosen such that none (or few) among the X_i actually transmit the infectious agent further onwards.

This estimator can be applied to HIV among men infected through sex with men in Denmark, Norway and Sweden. For these countries, estimation of the true absolute rates and the prevalence pool over time was carried out based on data from national registries of diagnosed HIV and AIDS cases [19–21]. The back-calculation models used were time-inhomogeneous Markov models, where new cases of HIV were assumed to follow a Poisson process. The models and results are described in Amundsen et al. [22], and the theoretical background is given by Aalen et al. [23]. Late reporting, possible under-reporting, and double reporting of HIV and AIDS are discussed and are found to have a minor influence on the results. Absolute rates were estimated from 1977 to 1995, and no positive values were assessed before 1977. Thus, the prevalence pool at time t could be calculated as the sum of HIV absolute rates from 1977 up to time t, minus an estimate of those persons who died or emigrated from 1977 up to time t. The number of deaths among HIV-positive persons before AIDS diagnosis was unknown, so population death rates among men aged 25–64 years were applied. Deaths among persons with AIDS were notifiable and available from the registries. Sensitivity studies were conducted regarding the effect on absolute rates of variation in deaths rates and emigration [22].

The uncertainty in the proposed estimator for R_a was estimated by a method described in Aalen et al., the 'fast' method [23]. The fast method is based on the maximum-likelihood estimates (MLE) of the absolute rates (X_i values) and their covariance matrix. One thousand estimations of R_a were simulated from a log-normal distribution with the MLE means and covariance structure for the X_i values. The absolute rates could be estimated separately for each half-year for the epidemic among men infected through sex with men in Denmark, Norway and Sweden, but with large uncertainties due to the low number of known cases. It was, therefore, necessary to join several subsequent intervals in the final model.

First, separate estimates were made for each half-year. Second, years with fairly similar estimates of absolute rates were joined to reduce the number of parameters and hence the uncertainty. This procedure induced fairly large jumps in absolute rates from one time-interval to the next in the peak period of the epidemic [22]. As a result, the absolute rate in this study was replaced by a moving average, with a band-width of 1½ years for Denmark, and 2½ years for Norway and Sweden. These band-widths were the narrowest which conserved the curve given by the originally estimated absolute rates over time, and also removed presumably uninformative variation in the estimates for R_a.

The average length of the infectious period in this study was set to 11 years. This represented the median incubation time from HIV infection to death for persons who were approximately 30 years old at infection during the period before the widespread use of highly active anti-retroviral treatment (HAART) [24]. Men infected through sex with men had a median age of 30–35 years at the time of diagnosis in all three countries, indicating a median age at the time of infection of 25–30 years in the Scandinavian countries. In the back-calculation models used to estimate absolute rates, the median time-period from infection to AIDS symptoms was 8·4 years up to and including 1987, while this period rose to 9·9 years in 1995. This reference data made a good model fit for Danish and Swedish data, while a rise to 9·3 years was applied to the Norwegian data, due to less use of AZT and other treatments during the period [22]. Time measures, from the onset of AIDS to time of death, were also added.

RESULTS

Among men infected through sex with men, the estimate of R_a varied substantially at the start of the epidemic, due to few cases. The estimated values of R_a are shown in the Figure from the point in time for each country at which the number of observations was large enough to give a stable curve, with a sufficiently narrow confidence interval (CI).

The estimates of actual HIV reproduction numbers were highest in 1981–1982 for Denmark and Norway, when each person would infect more than 15 secondary cases, on average. After 1981–1982, there followed 2–3 years when estimates of R_a decreased at the

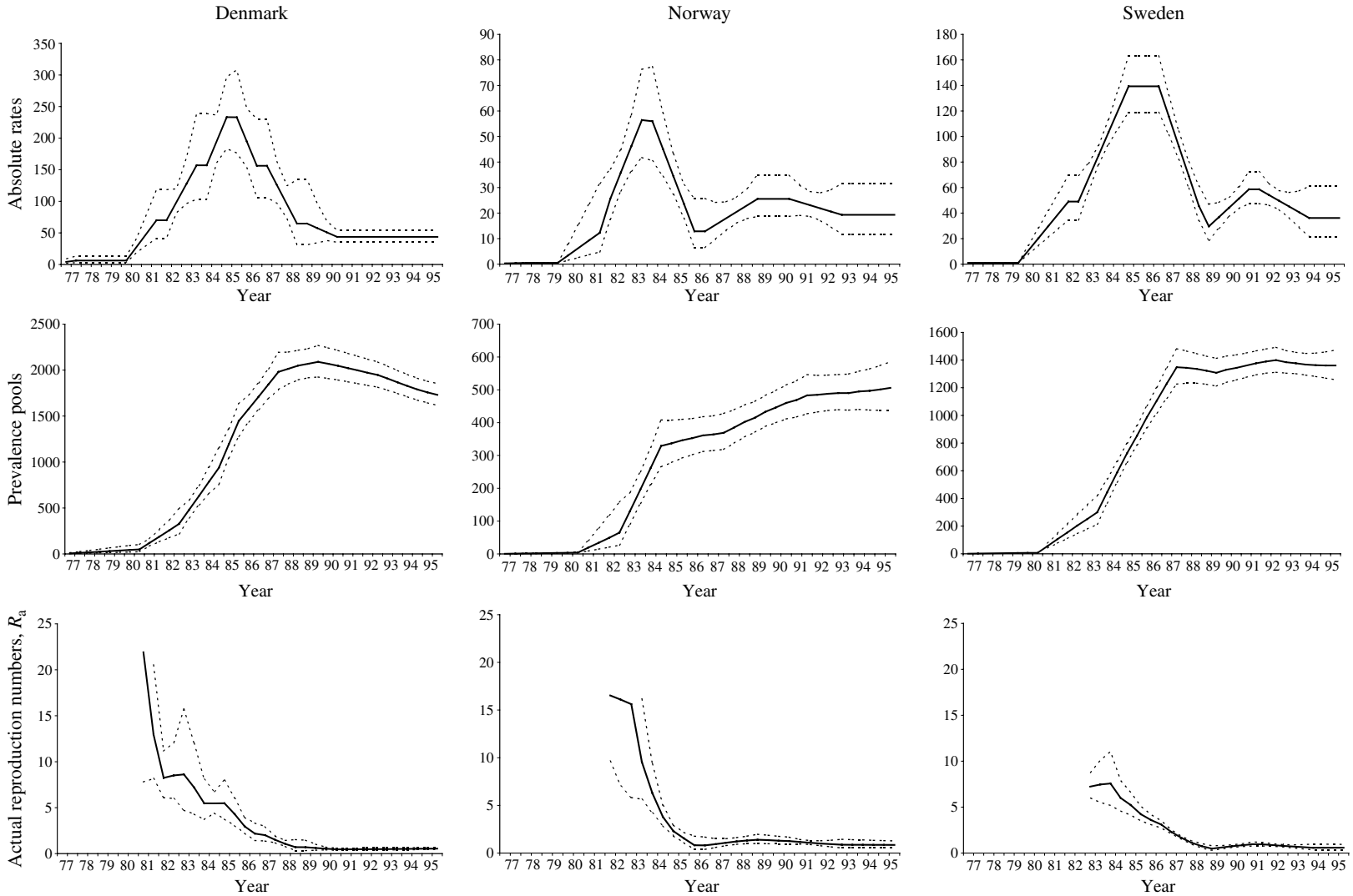

Fig. Estimates of HIV absolute rates, prevalence pools and the actual reproduction numbers among men having sex with men in Denmark, Norway and Sweden.

Table 1. *Periods with the estimator of R_a of <1 and significantly <1 for Denmark, Norway and Sweden**

	Denmark	Norway	Sweden
<1	1988.2–1995.2	1986.1–1987.1 1992.2–1995.2	1988.2–1995.2
Significantly <1	1990.1–1995.2	Never	1989.1–1990.1 1992.2–1995.2

* The notation 1988.2 means second half year of 1988 and 1990.1 means first half year of 1990.

same time as absolute rates and the prevalence pool in each country continued to increase.

The estimates of R_a became <1 from 1988 in Denmark and Sweden, and during 1986–1987 and 1992–1995 in Norway (see Table 1). R_a was significantly below this threshold for periods in Denmark and Sweden, but not in Norway.

By the end of the study period, the absolute rates had stabilized in all three countries. The estimate of R_a was 0·55 (95% CI 0·45–0·67) in Denmark, 0·85 (95% CI 0·57–1·27) in Norway and 0·58 (95% CI 0·36–0·95) in Sweden in 1995.

The prevalence pool decreased significantly in Denmark by the end of the study period, when R_a was most stable, to <1 after 1988. In Sweden, at this time, the prevalence pool stabilized. In Norway, where the R_a was never significantly <1, the prevalence pool still increased towards the end of the study period.

DISCUSSION

A measure for actual spread of an infectious disease in a population, the actual reproduction number, R_a, has been defined. An estimator for R_a has also been suggested, as constituting the new infected persons in an interval (the absolute rate), divided by the prevalence pool at the beginning of the interval, multiplied by the mean duration of the infectious period. As applied to the HIV epidemic among men infected through sex with men in the Scandinavian countries, the estimate of R_a declined from more than 15 secondary cases to approximately $\leqslant 1$ during the years, 1981–1995.

In this study application, the average length of the infectious period was set at 11 years. Until 1996, when HAART was introduced on a population-wide level in developed countries, the infectious period varied with the strength of an individual's immune system to fight the virus. Age at the time of infection is a good indicator of this strength [24]. Because a fairly similar mean age at diagnosis was identified in each of the three countries in our study, the length of the infectious period was not a critical aspect for comparison between the three countries, or over time. However, it is important with respect to inferences that may be drawn regarding the size of R_a. For comparisons of persons with different age distributions, different lengths of the infectious period should be used. After 1996, the length of the infectious period will vary more, depending also on access to HAART, the type of treatment, and personal adherence to treatment.

In this study application, the susceptible populations studied have not been isolated within their own countries. People travel due to employment and vacation. Some men get infected abroad having sex with men not in the prevalence pool in their home country. There may also be persons in the prevalence pool in a country that infect people from abroad who are not registered as having HIV in the same country, and do not live in the same country. The estimated value of R_a may, therefore, be biased.

It is difficult to state the direction of the bias for men infected through sex with men in the Scandinavian countries. It is possible to adjust the estimate for R_a down if information about newly diagnosed persons who were probably infected abroad can be deduced from the absolute rates. It is more difficult, however, to find information on the number of persons living abroad who were infected by persons in the country under study. The effect of this bias can be studied using sensitivity analysis, combining knowledge from behavioural studies and making reasonable assumptions. In addition, we note that misclassification of the mode of infection may also create bias.

Statistical variation in our study was assessed using the so called 'fast' method used by Aalen et al. [23]. The use of 1000 replications created stable distributions. The 'fast' method produced results similar to the more thorough and time-consuming bootstrap methods which could have been applied. A more direct distributional approach could also have been used since it is common to assume that the X_i values are independently Poisson-distributed and can be approximated by normal distributions; the distribution of the ratio between the means of two normally distributed variables, however, remains generally complex [25].

Regarding the confidence intervals for the estimates of R_a, these were wide during the early part of the epidemic: the upper 95% CI limit in the Figure was

Table 2. *HIV-diagnosed intravenous drug users in Latvia and Lithuania*

	1996	1997	1998	1999	2000	2001	2002	2003*
Absolute rates								
Latvia	—†	5	120	194	382	630	393	393
Lithuania	4	23	37	46	49	55	379	379
Prevalence pools‡								
Latvia	0	4·8	119·8	301	656	1235	1562	1877
Lithuania	4	25·9	60·4	102	145	192	548	890
Estimates of R_a§								
Latvia	—	—	277·3	17·7	13·9	10·6	3·5	2·8
Lithuania	—	63·3	15·7	8·4	5·3	4·2	21·7	7·6

* Figures for illustration only.
† Data prior to 1997 not available per transmission group.
‡ Sum of absolute rates minus 4% annual death estimate.
§ Assuming infectious period of 11 years.

62 for Denmark in 1981; for Norway, the values were 28, 37 and 42 for 1982–1983 (1982 – first half, 1982 – second half, and 1983 – first half, respectively). The confidence intervals for estimates of R_a in Sweden are given beginning in 1983.

Changes in R_a over time for HIV epidemics may be caused by several factors. In a situation with constant risk behaviour and a decreasing number of susceptible persons, the R_a would decline by necessity. Change can also follow from a change in sexual behavior at the individual level, or from a saturation of high activity groups and a diffusion of the virus into lower activity groups. Further, the average infectivity may change over time. All explanations for change in R_a – or one or more of them – may act at the same time. It is, therefore, not possible to identify directly the effect of HIV risk behaviour change in the Scandinavian countries from the change in R_a. The estimated distribution of persons in different stages of disease progression indicates only minor changes in average infectivity after the first initial period (data from back-calculation models is not shown). Although saturation in core groups may be a partial explanation for the reduction in R_a seen for the early part of the epidemic, changes in behaviour are probably a further reason [22]. Further, by using local knowledge, it is possible to get closer to probable reasons for the epidemic's change.

Additional and different conclusions can be reached from the estimates of R_a than are obtainable from the estimates of absolute rates and prevalence pools over time. First, in a period of increasing curves for the prevalence pools and the absolute rates, the number of secondary cases dropped dramatically for the same period. Note that the estimates of R_a for Denmark in 1981 were approximately 20 secondary cases, while by 1985, this figure had dropped to approximately six cases (see Fig.). The declining R_a values show that each infected person did, to a steadily lower degree, transmit the disease. This slowly decreasing rate of infection transmission is difficult to read from the two other measures, taken independently.

Second, additional conclusions are obtained based on the speed with which R_a decreases, including whether it reaches a value of <1. The epidemic in Norway did not reach this situation during the epidemic study period, with a value of R_a significantly <1, while the epidemics in Denmark and Sweden did. Due to similarities in the epidemic situation, this indicates that more prevention efforts were needed in Norway regarding HIV transmission among men having sex with men. Additionally, sustained values of <1 predict a declining prevalence pool. Sustained R_a values of <1 in Denmark were followed by a declining prevalence pool. The values of R_a will not necessarily decrease during an epidemic. If the agent (HIV, in our example) is spread to groups with higher risk activities, then R_a will increase again. Further, risk behaviour may contribute to maintain R_a values of >1 on a continuous basis.

A current example for HIV epidemics in a geographical area of interest, Latvia and Lithuania, is shown in Table 2, with data for newly diagnosed cases per year for intravenous drug users (IDUs) [26–28]. The absolute rates for 2003 were unknown at the time of writing, and were thus set as equal to the absolute rates for 2002, for illustration purposes only. Assume, for simplicity, that the rates shown in the first two

rows in Table 2 are the number of IDUs actually *infected* in the year given. In that case, the near-doubling of the absolute rates and prevalence pools in Latvia from 2000 to 2001 is synchronous with a reduction in the R_a estimate from 13·9 to 10·6. If the figures in the first row had been a good estimate of the true absolute rates of newly HIV-infected IDUs, Latvia would have seen good progress towards a situation with a declining epidemic. A possible reduction in absolute rates from 2002 to 2003 would improve the situation even more. However, further efforts are needed to reduce the R_a value to $\leqslant 1$.

The figures for Lithuania show a somewhat different development. A steady decrease in R_a was replaced by an alarming increase, from 2001 to 2002. A stable number of new cases in 2003, as in 2002, would then reduce the alarming R_a figure by 65%, from 21·7 to 7·6. A similar stable situation in absolute rates from 2002 to 2003 in Latvia is synchronous with only a 20% reduction in R_a. Knowledge of R_a, therefore, adds substantially different information than knowledge of the two referenced epidemic measures independently. We also note that because Table 2 is presenting diagnosed cases per year, the estimated figures for R_a are not correct for the true HIV epidemic development. More trustworthy estimates of R_a should be established and used in the surveillance of the epidemic, along traditional methodological lines. In other countries, the back calculation of absolute rates for HIV has been performed and estimates of R_a can thus be made [23, 29].

Both R_0 and R_e were defined locally in a short time-interval, or to a point in time, and are focused on studying the potential for epidemic outbreak. The measure, R_a, is focused on estimating the actual development that has taken place. Since the two measures are based on the same concept, it is tempting to study how the estimator (*) relates to R_e under various assumptions for the HIV epidemic. It can be shown, as below, and in the Appendix, that R_e and an expected value for the estimator (*) of R_a are equal given certain assumptions but unequal given others. Thus, it has been illustrated that the two situations, potential development and development that has taken place, should be conceptualized and managed separately.

First, a simple mathematical expression for R_0, which is relevant in homogenous populations with random mixing between partners in the case of the HIV epidemic is:

$$R_0 = c * \beta * D, \qquad (**)$$

where c is the average number of new partners per time unit, β is the average transmission rate per partner and D is the average length of the infectious period [3]. An expression for R_e in the same situation is:

$$R_e(t) = R_0 \frac{N(t) - P(t)}{N(t)}, \qquad (***)$$

at time t, where $N(t)$ is the population and $P(t)$ is the prevalence pool at time t [3]. Thus, $s(t) = (N(t) - P(t))/N(t)$ is the proportion of susceptibles at time t.

In this simple situation, assumed for (**) and (***) where c and β are constants, N_i is a given sample size at interval i, and P_{i-1} is assumed given, the expected number of new cases in interval i can be expressed as:

$$EX_i = P_{i-1} c s_{i-1} \beta, \qquad (****)$$

Here X_i is assumed to be the sum of 'successes' in n independent binomial trials where $n = P_{i-1} * c * s_{i-1}$ with the same probability, β, of transmitting the virus. Dividing (****) with P_{i-1} and multiplying with D, we get R_a equal to R_e, using (**) and (***). Thus, under the assumption of random mixing in homogenous populations, the expected value of (*), R_a and R_e are alike. The independence assumption is fulfilled when β is small, and this is true at least during the long non-symptomatic period after the primary infection.

In the Appendix, R_a and R_e are compared within a heterogeneous population, divided into groups with specified partner mixing. When an infection spreads in such a system, the number of newly infected will, after a while, reach a stationary distribution across the groups. The prevalence may still increase, but the distribution remains stable. The 'typical' infected referred to in the definitions of R_0 and R_e is an individual distributed according to this stationary distribution. Within these two groups, high risk and low risk, if the system stabilizes at 80% of newly infected coming from the high-risk group, the typical infected individual is 80% high risk and 20% low risk. This is clearly a very theoretical concept.

In contrast, we can see from examining the definition of R_a that it will reflect the average behaviour of the newly infected at each time-interval. We show, in the Appendix, that the mean of (*) and R_e will be equal if the following conditions are met: (1) there is the same average infectiousness in all groups; (2) there is no transition of infections between the groups; and (3) the newly infected have reached a stationary distribution. Note that we have not attempted to show equality in the absence of (1) and (2)

above, but the equality may hold also under weaker conditions. However, condition (3) is essential.

The time needed to reach a stationary distribution depends on how strongly the groups are linked (strong partner mixing), and on how far from stationary the system is initially (for example, whether the first infection occurred in the high- or the low-risk group). If the distribution of newly infected over the groups is far from the stationary distribution, the mean of (*) and R_e will be different. R_e will measure the potential spread at this point in time of the epidemic, whereas R_a will measure the actual spread at this point in time of the epidemic.

Another path for development is to find other estimators for R_a. Instead of looking at the potential if the situation in a time-interval is assumed to hold for the whole infectious period, as in (*), an historical estimator is suggested, using the data available here (X_j and P_{j-1}). The expression is:

$$\hat{R}_h(i) = \sum_{j=i}^{i+m} \frac{X_j}{P_{j-1}}, \qquad (*****)$$

where m is the number of intervals in the infectious period D, and the interval i is the half-year of infection. The R_h (first half-year of 1983) shows how many secondary cases one person produced, on average, in the 11 years from 1983 to 1994. Applied to the data here, this is a much smaller figure than the value of R_a in the first half of 1983 because conditions have changed during the period. But $R_h(i)$ can only be calculated for a short calendar period, i, covering the period 1982–1984 in our data. Before 1982, stability was low, and, after 1995, this study ended.

We also note, however, that the historical estimator suggested in (*****) is biased if risk activity and other factors regulating HIV transmission vary with time, since infection caused by newly infected persons entering the prevalence pool over calendar time will contribute to the estimate. If calendar time is a more important factor for HIV transmission (*****) may be a useful estimator for the average actual number of persons infected by one infected person over the infectious period.

To our knowledge, no material on the ratio between empirical absolute rates (X_i) and the prevalence pools of HIV (P_{i-1}) over time has been published [9]. However, in the case of tuberculosis, a fairly similar approach has been applied [30, 31]. A transmission index was defined as the average total number of cases of tuberculosis attributable, directly or indirectly, to recent transmission from a single source case. The ratio, X_i/P_{i-1} in (*), can be seen as such a transmission index, where it is not possible to distinguish between secondary cases and transmissions further on. Using short time-intervals in the analysis, such further transmissions from the secondary cases will be scarce. It seems more meaningful to find estimators for R_a, instead of utilizing a general transmission index, thus staying with the more conceptually fruitful ideas of reproduction numbers.

The suggestions made here may also be useful in the analysis of other infectious diseases, although further work should be carried out to establish the usefulness of both the concept and its estimators in different epidemic situations. In addition, better designs for particular studies should provide alternative ways to estimate R_a [32]. When a new infectious agent occurs or a known one re-occurs, one should try to gather data to estimate R_a from (*) to see if the number increases or decreases over time, and how quickly it is changing. The duration of infectiousness, D, may, for example, not be known. However, the trend in the ratios X_i/P_{i-1} for known intervals is also useful. Even a rough idea of R_a would lead to a better understanding of the epidemic and the needs of present and future efforts for disease prevention and control. The interpretation of the value of R_a and the changes in that value from one interval to the next should be carried out by researchers with a good knowledge of the infectious agent, together with persons knowing how the data was established.

If incidence rates and the prevalence over time are available from other types of data, models or estimation techniques, the incidence rates divided by the prevalence at a given point in time for low-incidence diseases would approximate the value of the absolute rate divided by the prevalence pool. This could be a useful check. Additionally, new methods of identifying and tracing infectious agents are evolving, in which case R_a may, more often, be fully or partly estimated directly, as in Potterat et al. [6].

In conclusion, actual reproduction numbers for HIV epidemics that have taken place can be estimated from incidence and prevalence information, and this study application shows interesting features. Estimates of actual reproduction numbers can provide us with a new understanding of the dynamics of epidemics, in addition to what can be learned from the prevalence and incidence information. Further work is necessary to establish the usefulness of both the concept and its estimators for different infectious diseases in various epidemic situations.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support which made this work possible. This work was financed, in part, by the Norwegian Ministry of Health and Social Affairs for 1996–2001, the University of Oslo for 1998–2001, the Norwegian Institute of Public Health for 2001; and the Norwegian Institute for Alcohol and Drug Research (SIRUS) for 2002–2003.

APPENDIX

Comparison of the mean of the estimator (*) for the actual reproduction number, R_a, with the effective reproduction number, R_e, in a system of interacting groups

Let $P(t)$ be a vector of the number of infected in each group at time t in a system of interacting groups. The change in the number of infected can be described by a system of differential equations:

$$P'(t) = \mathbf{M}(t) \cdot P(t),$$

The matrix $\mathbf{M}(t)$ contains elements that describes the new infections, as well as rates for transitions between groups and out of the system. In contrast to the use of Markov models in the same situation, the absorbing state is not included in the system. The matrix $\mathbf{M}(t)$ can be split in two so that $\mathbf{M}(t) = \mathbf{M}_1(t) + \mathbf{M}_2$, with $\mathbf{M}_1(t)$ generating the new infections, and $\mathbf{M}_2$ containing the transitions. A simple example with two groups is:

$$\mathbf{M}_1(t) = \begin{bmatrix} \frac{n_1 - p_1(t)}{n_1}\pi_{11}c_1\beta & \frac{n_1 - p_1(t)}{n_2}\pi_{12}c_1\beta \\ \frac{n_2 - p_2(t)}{n_1}\pi_{21}c_2\beta & \frac{n_2 - p_2(t)}{n_2}\pi_{22}c_2\beta \end{bmatrix}$$

$$\mathbf{M}_2 = \begin{bmatrix} -1/d & 0 \\ 0 & -1/d \end{bmatrix}, \quad P(t) = \begin{bmatrix} p_1(t) \\ p_2(t) \end{bmatrix},$$

where n_i is the number of subjects, $p_i(t)$ the number of infected and c_i is the contact rate in group i. π_{ij} is the proportion of group i partners that come from group j, β is the transmission rate per contact, and d is the duration of infectiousness. In this simple example, there is partner mixing but no transitions between the groups (see Stigum et al. for a more elaborate example [18]).

Based on the transition matrix $\mathbf{M}_2$, we can find a matrix $\mathbf{L}$ giving the expected length of stay in each of the groups or states. The matrix product $\mathbf{L} \cdot \mathbf{M}_1(t) = \boldsymbol{G}(t)$ gives the expected number of new cases produced in each state during the average infectious period, and is called the next generation matrix [11–13].

The basic reproduction number R_0 is defined as the largest eigenvalue of $\boldsymbol{G}(0)$ [11–13]. We suggest that the effective reproduction number $R_e(t)$ be defined as the largest eigenvalue of $\boldsymbol{G}(t)$. This definition of $R_e(t)$ gives us the number of new cases produced by a typical infected subject during the infected period, in a population with a number of susceptibles as at time t. The phrase 'typical infected' means that the newly infected subjects are distributed according to the stationary distribution of the system defined at time t.

Let the dominant eigenvalue of $\boldsymbol{G}(t)$ be called $\lambda(t)$, and the corresponding eigenvector be $P_\lambda(t)$.

An estimator for the actual reproductive number $R_a(t)$ is suggested in this paper to be the number of new cases per existing cases times the duration of infectiousness. Let v be a line vector of ones. Then, the mean of the estimator for $R_a(t)$ can be written:

$$R_a(t) \equiv \frac{\text{new cases}}{\text{existing cases}}\text{duration} = \frac{v \cdot \mathbf{M}_1(t) \cdot P(t)}{v \cdot P(t)}\, d.$$

$\mathbf{M}_1(t)$ times $P(t)$ gives a vector of the new cases, multiplication with v will give us the sum of these. If all groups have the same average duration of infectiousness and there are no transitions between the groups, the $\mathbf{L}$ matrix giving expected length of stay will simply be a diagonal matrix with d on the diagonal. Then $\mathbf{L} = \mathbf{I}d$ where $\mathbf{I}$ is the identity matrix. Therefore, the following holds:

$$R_a(t) = \frac{v \cdot \mathbf{M}_1(t) \cdot P(t)}{v \cdot P(t)}\, d = \frac{v \cdot \mathbf{I} \cdot \mathbf{M}_1(t) \cdot P(t)}{v \cdot P(t)}\, d$$

$$= \frac{v \cdot \mathbf{L} \cdot \mathbf{M}_1(t) \cdot P(t)}{v \cdot P(t)} = \frac{v \cdot \boldsymbol{G}(t) \cdot P(t)}{v \cdot P(t)}.$$

If the distribution of cases $P(t)$ over the groups has become stationary, $P(t)$ will be proportional to the eigenvector of the dominant eigenvalue $P_\lambda(t)$. Given this type of stationary distribution, we have:

$$R_a(t) = \frac{v \cdot \boldsymbol{G}(t) \cdot P(t)}{v \cdot P(t)} = \frac{v \cdot \lambda(t) \cdot P_\lambda(t)}{v \cdot P_\lambda(t)} = \lambda(t) \equiv R_e(t).$$

We have shown that the mean for our estimator of the actual reproductive number will be equal to the effective reproductive number as defined above if all groups have the same average infectiousness, there are no transitions between the groups, and the system has reached a stationary distribution of infected persons over the groups. R_e measures the potential for spread.

If there are weak links between some of the groups – for instance, almost no partner mixing between two groups (i.e. the system is almost reducible), then the distribution of infected subjects will not reach a stationary distribution for a long time, but will depend on the initial conditions. In this case, R_e will be different from R_a, R_a measuring the actual transmission taking place.

REFERENCES

1. Rothman K, Greenland S. Modern epidemiology. Philadelphia: Lippincott-Raven, 1998: 36, 42.
2. Anderson RM, May RM. Infectious diseases of humans. Dynamics and control. Oxford: Oxford University Press, 1991: 17.
3. Brookmeyer R, Gail MH. AIDS epidemiology: a quantitative approach. New York: Oxford University Press, 1994: 234, 239.
4. Halloran ME. Concepts of transmision and dynamics. In: Thomas JC, Weber DJ, eds. Epidemiological methods for the study of infectious diseases. New York: Oxford University Press, 2001: 63–64.
5. Hethcote HW. The mathematics of infectious diseases. Siam Review 2000; **42**: 599–653.
6. Potterat JJ, Zimmerman-Rogers H, Muth SQ, et al. Chlamydia transmission: concurrency, reproduction number, and the epidemic trajectory. Am J Epidemiol 1999; **150**: 1331–1339.
7. Haydon DT, Chase-Topping M, Shaw DJ, et al. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. Proc R Soc Lond B Biol Sci 2003; **270**: 121–127.
8. Borgdorff MW, Nagelkerke N, van Soolingen D, et al. Analysis of tuberculosis transmission between nationalities in the Netherlands in the period 1993–1995 using DNA fingerprinting. Am J Epidemiol 1998; **147**: 187–195.
9. Pinkerton S, Abramson P, Kalichman S, et al. Secondary HIV transmission rates in a mixed-gender sample. Int J STD AIDS 2000; **11**: 38–44.
10. Chick SE, Adams AL, Koopman JS. Analysis and simulation of a stochastic, discrete-individual model of STD transmission with partnership concurrency. Math Biosci 2000; **166**: 45–68.
11. Diekmann O, Heesterbeek JAP, Metz JAJ. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogenous populations. J Math Biol 1990; **28**: 365–382.
12. Diekmann O, Dietz K, Heesterbeek JAP. The basic reproduction ratio for sexually transmitted diseases. 1. Theoretical considerations. Math Biosci 1991; **107**: 325–339.
13. Dietz K. The estimation of the basis reproduction number for the infectious diseases. Stat Methods Med Res 1993; **2**: 23–41.
14. Farrington CP, Kanaan MN, Gay NJ. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. J R Stat Soc Ser C – App Stat 2001; **50**: 251–283.
15. Farrington CP, Whitaker HJ. Estimation of effective reproduction numbers for infectious diseases using serological survey data. Biostat 2003; **4**: 621–632.
16. Farrington CP. On vaccine efficacy and reproduction numbers. Math Biosci 2003; **185**: 89–109.
17. Longini IM, Halloran ME, Nizam A. Model-based estimation of vaccine effects from community vaccine trials. Stat Med 2002; **21**: 481–495.
18. Stigum H, Magnus P, Bakketeig LS. Effect of changing partnership formation rates on the spread of sexually transmitted diseases and human immunodeficiency virus. Am J Epidemiol 1997; **145**: 644–652.
19. Smith E, Rix BA, Melbye M. Mandatory anonymous HIV surveillance in Denmark: the first results of a new system [see comments]. Am J Public Health 1994; **84**: 1929–1932.
20. Arneborn M, Giesecke J. HIV and AIDS statistics from the Swedish Institute for Infectious Disease Control [in Swedish]. In: Aktuell information från Smittskyddsinstitutet [Current information from the Swedish Institute for Infectious Disease Control]. Stockholm, 1997; **1**: 7–8.
21. Aavitsland P, Nilsen Ø, Lystad A. Anonymous reporting of HIV infection: an evaluation of the HIV/AIDS surveillance system in Norway 1983–2000. Eur J Epidemiol 2001; **17**: 479–489.
22. Amundsen EJ, Aalen OO, Stigum H, et al. Back-calculation based on HIV and AIDS registers in Denmark, Norway and Sweden 1977–95 among homosexual men: estimation of absolute rates, incidence rates and prevalence of HIV. J Epidemiol Biostat 2000; **5**: 233–243.
23. Aalen OO, Farewell VT, de Angelis D, et al. A Markov model for the AIDS incubation time including the effect of HIV-diagnosis and treatment: application to AIDS prediction in England and Wales. Stat Med 1997; **16**: 2191–2210.
24. Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action. Concerted Action on SeroConversion to AIDS and Death in Europe. Lancet 2000; **355**: 1131–1137.
25. Johnson N, Kotz S, Balakrishnan N. Continous univariate distributions. New York: Wiley, 1994: 327–328.
26. HIV/AIDS Surveillance in Europe: end-year report 2000; no. 64: 31. St Maurice, France, European Centre for the Epidemiological Monitoring of AIDS, 2001.
27. HIV/AIDS Surveillance in Europe: end-year report 2002; no. 68: 19. St Maurice, France, European Centre for the Epidemiological Monitoring of AIDS, 2003.
28. HIV/AIDS Surveillance in Europe: mid-year report 2003; no. 69: 19. St Maurice, France: European

Centre for the Epidemiological Monitoring of AIDS, 2003.
29. Back-calculation estimates of HIV cumulative incidence and prevalence to 31 December 1993 and predicted annual numbers of AIDS cases to 1998 among adults/adolescents in Europe; no. 44: 55–68. St Maurice, France, European Centre for the Epidemiological Monitoring of AIDS, 1994.
30. Van Soolingen D, de Haas PE, Kremer K, et al. Molecular epidemiology of tuberculosis in a low incidence country: a nation-wide study of transmission of tuberculosis between immigrants and the native population in the Netherlands. In: Van Soolingen D, ed. Use of DNA fingerprinting in the epidemiology of tuberculosis. Utrecht, State University of Utrecht, 1996: 175–195.
31. Van Soolingen D, Lambregts-van Weezenbeek C, de Haas PE, et al. Transmission of susceptible and resistant Mycobacterium tuberculosis strains in The Netherlands, 1993–95, investigated using DNA fingerprinting [in Dutch]. Ned Tijdschr Geneeskd 1996; **140**: 2286–2289.
32. Becker NG, Britton T. Design issues for studies of infectious diseases. J Stat Plan Inf 2001; **96**: 41–66.