Indirect methods for estimating prevalent HIV infections: adults in England and Wales at the end of 1993

G. HUGHES*, K. PORTER and O. N. GILL

PHLS AIDS Centre, Communicable Disease Surveillance Centre, 61 Colindale Avenue, London, UK NW9 5EQ

(Accepted 16 December 1997)

SUMMARY

Two indirect methods were used to estimate the point prevalence of HIV infection in England and Wales at the end of 1993 using data on diagnosed HIV infections, AIDS cases, HIV-related deaths and HIV testing behaviour from unlinked anonymous surveys. The methods estimated the proportion of all prevalent HIV infections that diagnosed infections represented. Most of those exposed to HIV infection through injecting drug use or sexual intercourse between men had had their infections diagnosed compared to less than half of those exposed through heterosexual intercourse. The total estimated number of prevalent infections was 22,350 for the diagnosis interval method and 20,540 for the test history method, and about 56–57% of these were in homosexual/bisexual men. These indirect methods are cheap and simple applications of surveillance data which provide estimates that compare favourably with those produced by more complex methods.

INTRODUCTION

Reports of infection with the human immunodeficiency virus (HIV) in individuals who voluntarily present themselves for an HIV test provide a valuable insight into the nature of the HIV epidemic. Taken on their own, however, such reports present a biased picture, reflecting infections in those who choose to be tested. HIV testing behaviour is known to vary considerably across population subgroups [1] and is probably directed by an individual's perception of risk, illness and concern about confidentiality. Methods which estimate the number of prevalent HIV infections, and how these are distributed within the population, are essential for understanding the full extent of the epidemic. Furthermore, independent estimates of the range of prevalent HIV infections can be used to refine models which estimate future AIDS cases [2, 3].

Individuals who are infected with HIV can be stratified into two groups: (1) those who have had their infection diagnosed by a voluntary HIV test and (2) those who have not had a voluntary HIV test and are unaware of their infection status.

In this paper, two indirect approaches to estimating the prevalent number of HIV-infected individuals who have not had a voluntary and confidential test, are described and earlier prevalence estimates [4] updated. The methods estimate the proportion of all infections which have been diagnosed [5–7] and are based on a national reporting system for diagnosed infections which include sufficient information to allow duplicate reports to be eliminated (a situation which exists in few countries). The need for population-based surveillance of HIV and AIDS diagnoses in all states of the USA to help provide better estimates of the number of HIV-infected persons has recently been highlighted [8].

The first method, the diagnosis interval method, is based on the proportion of persons recently diagnosed with AIDS whose first positive HIV test was probably before any illness associated with their AIDS diagnosis. This proportion is assumed to indicate the
proportion of all persons currently infected with HIV without AIDS who had a voluntary HIV test. The second method, termed the test history method, uses limited information on HIV testing behaviour from the unlinked anonymous surveys [1]. The proportion of infected individuals in a given survey period who had not had a voluntary and confidential HIV test is assumed to be equivalent to the proportion of undiagnosed HIV infections.

A direct method for estimating prevalent HIV infections has also been described [9–11] and involves combining HIV prevalence in defined behavioural groups as measured by unlinked anonymous sera-surveys [1] with population sizes estimated using data from the National Survey of Sexual Attitudes and Lifestyles [12] and the Office for National Statistics [13].

Both indirect methods were used, along with the direct method, to produce preferred estimates of prevalent HIV infections in behavioural subgroups in England and Wales at the end of 1988 [7], 1991 [4] and 1993 [9]. The 1991 [4] and 1993 [9] estimates were incorporated into models for projecting the number of AIDS cases.

**METHODS**

**Estimating prevalent diagnosed HIV infections**

Reports of antibody positive, voluntary and confidential HIV tests, AIDS diagnoses and deaths in HIV-infected people in England and Wales to the Public Health Laboratory Service (PHLS) AIDS Centre at the Communicable Disease Surveillance Centre (CDSC) were used in the calculations. Data from reports to the end of June 1995 for adults (> 14 years old at positive test) who were presumed to reside in the UK, were used. The reports were adjusted for undetected duplication and under-reporting (or under-ascertainment). Under-reporting comprises delayed reporting (reports which are received late) and non-reporting (reports which are never received).

The extent of under-reporting of HIV diagnoses has been estimated from the number of AIDS case reports which did not have a matched HIV infection report, since almost all AIDS cases reported in the UK since 1985 will have had an HIV test to confirm the diagnosis. Using this method the completeness of national reporting of HIV infections has been estimated to be between 80 and 84% and, with no evidence to the contrary, assumed to be similar across exposure categories (J. Mortimer, personal communication). Based on this estimate, cumulative reported HIV infections were multiplied by 1.18 to allow for 15% under-reporting and give a best estimate of cumulative diagnosed HIV infections. A range for the best estimate of diagnosed infections was given by multiplying cumulative infection reports by 1.25, a plausible upper limit for under-reporting of 20%, and by 0.95 to allow for 5% undetected duplicate reporting.

Cases of AIDS were adjusted for reporting delay following the method of de Angelis and Gilks [14] and for non-reporting by multiplying by 1.15 (13%) [15]. Deaths from AIDS were corrected for delayed ascertainment [9] and for non-ascertainment by multiplying by 1.15 (13%) [15]. All calculations used the adjusted figures and numbers were rounded to the nearest 10.

An estimate of the number of prevalent diagnosed HIV infections without AIDS (Y, Table 1) in England and Wales at the end of 1993 was calculated by subtracting the number of deaths in HIV infected people (including those without AIDS) and prevalent AIDS cases at the end of 1993 from the cumulative number of diagnosed HIV infections at the end of the 1993.

**Estimating total prevalent HIV infections**

**Diagnosis interval method**

Individuals with AIDS can be divided into two groups, comprising those who had an HIV test well in advance to their AIDS diagnosis and those who were not known to be infected until they developed an illness followed shortly by the diagnosis of an AIDS indicator disease. For the diagnosis interval method it is assumed that the proportion of recent incident AIDS diagnoses, in whom the interval between their first HIV positive test and their AIDS diagnosis (the ‘diagnosis interval’) is relatively long, represents the proportion of all persons infected with HIV (who are alive and without AIDS) who have had a voluntary HIV test. Certain AIDS cases were omitted from calculations estimating diagnosis intervals: (1) those whose year of first positive test and/or AIDS diagnosis was unknown (7% of cases) and (2) those with a diagnosis interval of less than 2 years and whose month of first positive test and/or AIDS diagnosis was unknown (1% of remaining cases).

The proportion of AIDS cases (in homosexual and bisexual men in the Thames regions) with intervals of over 3, over 9 and over 12 months between first
Indirect methods to estimate HIV prevalence

167

positive HIV test and AIDS diagnosis has risen gradually and in parallel since 1989 (data not shown). For the present calculations, the most suitable cut-off point for a long diagnosis interval was judged to be over 9 months to exclude those who sought an HIV test following an illness which led to an AIDS diagnosis relatively quickly, and to minimize exclusion of persons with an HIV diagnosis who were asymptomatic for AIDS.

The groups were defined as \( N_1 \), the number of AIDS cases diagnosed in 1992 and 1993 with an interval of 9 months or less between HIV diagnosis and AIDS diagnosis (short diagnosis interval), and \( N_2 \), the number of AIDS cases diagnosed in 1992 and 1993 with an interval of 10 months or longer between HIV diagnosis and AIDS diagnosis (long diagnosis interval). The proportion of AIDS cases diagnosed in 1992 and 1993 who had a long diagnosis interval (\( p \)) is therefore given by:

\[
p = \frac{N_2}{N_1 + N_2}.
\]

If it is assumed that progression to AIDS is equal for individuals with long and short diagnosis intervals, then the total number of prevalent undiagnosed HIV infections can be estimated by

\[
X = \frac{Y}{p} - Y,
\]

where \( X \) is the number of prevalent undiagnosed HIV infections and \( Y \) is the number of prevalent diagnosed HIV infections without AIDS at the end of 1993.

Total prevalent HIV infections at the end of 1993 is therefore given by

\[
X + Y + Z,
\]

where \( Z \) is the total number of prevalent AIDS cases at the end of 1993.

Individuals were grouped into the following exposure categories.

1. Homosexual and bisexual males (including a few who had injected drugs):
   - (a) Thames regions;
   - (b) Rest of England and Wales.
2. Injecting drug users (IDUs).
3. Heterosexuals (excluding those who had injected drugs or received blood, tissue or blood products):
   - (a) exposure abroad;
   - (b) exposure not known to be abroad (included those with a ‘high risk’ partner, such as a bisexual man, a haemophiliac or an injecting drug user).
4. Blood, tissue or blood product recipients.

Stages of the calculation are shown for homosexual and bisexual men in Table 1.

**Test history method**

The test history method is based on the proportion of infected individuals within a given population who have had their infection confirmed by a voluntary and confidential HIV test. To calculate the proportion of individuals in specific exposure groups who are infected with HIV and who have had a voluntary HIV test data were collected from three sources:

1. The unlinked anonymous serosurvey of genitourinary medicine (GUM) clinic attenders [1].
2. The unlinked anonymous survey of injecting drug users co-ordinated by the PHLS AIDS Centre [1].
3. The survey of injecting drug users co-ordinated by the Centre for Research on Drugs and Health Behaviour [16].

Data from the unlinked anonymous serosurvey of GUM clinic attenders [1] provided an estimate of the proportion of all infected individuals presenting for syphilis serology at GUM clinics who had been diagnosed infected with HIV by a voluntary and confidential test. These data are supplied by the attending physician who either asks the patients for this information or obtains it from the case notes. The testing history of individuals infected with HIV was estimated from data collected in 1993 from five clinics in London and eight clinics outside London. Clinic attenders known to have AIDS were excluded.

Data for 1992 and 1993 from the surveys of injecting drug users [1, 16] were combined to give an overall estimate of the proportion of HIV-infected injecting drug users who had had their infections diagnosed by a positive voluntary HIV test. The information was obtained by asking injectors to complete a questionnaire. The survey conducted at the PHLS AIDS Centre [1] did not record the test result of those infected people who had previously had an HIV test. This is relevant because a small number of newly infected individuals may previously have had an HIV test with a negative result: such individuals would not have had their infections diagnosed. The proportion of infected people who had previously tested negative was estimated using equivalent data from the survey co-ordinated by the Centre for Research on Drugs and Health Behaviour [16]. These
Table 1. Estimate of prevalent HIV infections in homosexual and bisexual men at the end of 1993 in England and Wales using the diagnosis interval method

<table>
<thead>
<tr>
<th>Exposure category</th>
<th>Diagnosed HIV infections reported (range)*</th>
<th>Best estimate of diagnosed HIV infections‡</th>
<th>Known deaths in HIV infected people</th>
<th>Best estimate of deaths in HIV infected persons‡</th>
<th>Prevalent AIDS cases§ (Z)</th>
<th>Prevalent diagnosed HIV infections without AIDS (range) (Y)</th>
<th>Proportion of AIDS cases with a ‘diagnosis interval’ &gt; 9 months (95% CI) (p)¶</th>
<th>Prevalent undiagnosed HIV infections (range) (X)¶</th>
<th>Total prevalent HIV infections (range) (X + Y + Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo/bisexual males Thames regions</td>
<td>9540 (9060–11330)</td>
<td>11260</td>
<td>3850</td>
<td>4390</td>
<td>1700</td>
<td>1800</td>
<td>5070 (2870–5140)</td>
<td>0.69 (0.67–0.72)</td>
<td>2280 (1290–2310)</td>
</tr>
<tr>
<td>Rest of England and Wales</td>
<td>3050 (2900–3630)</td>
<td>3600</td>
<td>1170</td>
<td>1330</td>
<td>430</td>
<td>460</td>
<td>1810 (1110–1840)</td>
<td>0.61 (0.57–0.66)</td>
<td>1160 (710–1180)</td>
</tr>
</tbody>
</table>

* Lower limit calculated by multiplying by 0.95 (to allow for duplicate reporting) and upper limit by multiplying by 1.25 (to give a plausible maximum for under-reporting).
† Cumulative reported HIV infections multiplied by 1.18 (15%) to adjust for under-reporting (see text).
‡ Deaths from AIDS corrected for ascertainment delay as in Day [8] and for non-ascertainment by multiplying by 1.15 [14].
§ AIDS cases adjusted for reporting delay following de Angelis and Gilks [13] and for non-reporting by multiplying by 1.15 [14].
¶ Calculated from AIDS cases diagnosed in 1992 and 1993, where \( p = N_x/N_1 + N_v \). \( N_1 \) are cases with an interval of 9 months or less between HIV diagnosis and AIDS diagnosis, and \( N_v \) are cases for which this interval is 10 months or longer.
¶¶ Assuming \( Y/(Y + X) = p \), then \( X = (Y/p) - Y \).
Indirect methods to estimate HIV prevalence

Fig. 1. The proportion of AIDS cases diagnosed in 1992 and 1993 for which the interval between first positive HIV test and AIDS diagnosis (diagnosis interval) was greater than 9 months and the proportion of HIV-infected persons in unlinked anonymous (UA) surveys whose infections were diagnosed by a voluntary and confidential HIV test, stratified by exposure (see text). The UA surveys were: the unlinked anonymous serosurvey of genitourinary medicine clinic (GUM) attenders [1], the voluntary unlinked anonymous survey of injecting drug users (IDU) co-ordinated by the PHLS AIDS Centre [1] and the survey of IDUs co-ordinated by the Centre for Research on Drugs and Health Behaviour [15]. Confidence intervals were calculated using the exact method.

individuals were assumed not to have had their infections diagnosed in order to prevent over-estimation of the proportion of diagnosed HIV infections. It was assumed that few IDUs diagnosed with AIDS were likely to be attending drug centres. For a minority of cases HIV testing history was not known; such cases were omitted from the calculations.

If it is assumed that the proportion of infected individuals in the unlinked anonymous surveys who were tested \((p_x)\) is equivalent to the proportion of all persons infected with HIV without AIDS in England and Wales who have had a voluntary and confidential HIV test, then the number of undiagnosed infections without AIDS \((X_x)\) can be given by

\[
X_x = \frac{Y}{p_x} - Y,
\]

and the total number of prevalent HIV infections by

\[
X_x + Y + Z,
\]

where \(Y\) and \(Z\) are defined as for the diagnosis interval method.

Estimates of the number of prevalent HIV infections were calculated for three exposure categories, (i) homosexual and bisexual men (including some who had ever injected drugs), (ii) heterosexual men and women (excluding those who had ever injected drugs) and (iii) injecting drug users. The preferred estimate for injecting drug users was taken as the arithmetic mean of the estimates calculated from the GUM clinic attenders survey [1] and the two surveys of injecting drug users [1, 16].

As the diagnosis interval method is the only method which estimates HIV prevalence in those exposed to infection through blood or tissue transfer or blood factor treatment, this estimate was used throughout to allow total estimates of HIV infection prevalence to be made.

RESULTS

The majority of those diagnosed with AIDS during 1992 and 1993 who were exposed to infection through injecting drug use or through sexual intercourse between men had received a positive voluntary and confidential HIV test at least 10 months prior to their diagnosis of AIDS (Fig. 1). Similarly, of those in the unlinked anonymous survey [1, 16] who were infected with HIV and who were also in these exposure categories, the majority had had their HIV infections diagnosed (Fig. 1). In contrast, less than half of those exposed through heterosexual intercourse who were diagnosed with AIDS in 1992 or 1993 had received an HIV diagnosis 10 or more months prior to their AIDS diagnosis (Fig. 1). The unlinked anonymous survey of GUM clinic attenders [1] also indicates that of those infected with HIV who were exposed to infection through heterosexual intercourse, less than 40% had had their infections diagnosed (Fig. 1).
Table 2. Estimates of prevalent HIV infections (with ranges) in England and Wales at the end of 1993 using two indirect methods and a direct method*

<table>
<thead>
<tr>
<th>Exposure category†</th>
<th>Diagnosis interval</th>
<th>Test history</th>
<th>Direct*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo/bisexual males</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thames regions</td>
<td>9150 (5960–9250)</td>
<td>8650 (5680–8750)</td>
<td>7800 (7000–8600)</td>
</tr>
<tr>
<td>Rest of England and Wales</td>
<td>3430 (2280–3480)</td>
<td>3120 (2090–3170)</td>
<td>4900 (4400–5400)</td>
</tr>
<tr>
<td>Injecting drug users</td>
<td>1880 (1420–1900)</td>
<td>1770 (1350–1800)</td>
<td>2500 (2300–2800)</td>
</tr>
<tr>
<td>Heterosexual exposure</td>
<td>6290 (4600–6390)</td>
<td>7000 (6300–7700)</td>
<td></td>
</tr>
<tr>
<td>Exposure abroad</td>
<td>5960 (4340–6060)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure not known to be abroad</td>
<td>1200 (890–1240)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood, tissue or blood factor recipients‡</td>
<td>710 (440–720)</td>
<td>710 (440–720)</td>
<td>710 (440–720)</td>
</tr>
<tr>
<td>Total§</td>
<td>22350 (17560–22540)</td>
<td>20540 (16110–20720)</td>
<td>22910 (21400–24420)</td>
</tr>
</tbody>
</table>

* See reference [9].
† Fuller description of exposure categories in text.
‡ Diagnosis interval method estimate used throughout.
§ Range for the total:
  Lower range = total central estimate – \(\sqrt{\Sigma (central\ estimate – lower\ estimate)^2}\).
  Upper range = total central estimate + \(\sqrt{\Sigma (upper\ estimate – central\ estimate)^2}\).

Final estimates of prevalent HIV infections in England and Wales at the end of 1993 using the diagnosis interval method and the test history method are presented with estimates produced using the direct method [9] in Table 2. Overall, the estimated number of prevalent HIV infections in adults in England and Wales at the end of 1993 was 22350 and 20540 for the diagnosis interval and test history methods, respectively.

The number of estimated infections based on the diagnosis interval method exceeded those calculated using the test history method for each exposure category. For the diagnosis interval and test history methods, respectively, estimates were 12580 and 11700 (56 and 57% of estimated infections) for those presumed infected through homosexual intercourse, 1880 and 1770 (8 and 9%) for those exposed through injecting drug use and 7180 and 6290 (32 and 31%) for those exposed through heterosexual intercourse. The majority of those exposed through sexual intercourse between men were in the Thames regions (73 and 75% of infections in this exposure group for the diagnosis and test history methods, respectively). About 83% of those exposed through heterosexual intercourse (5960 of the 7180 estimated using the diagnosis interval method) were probably exposed abroad. The diagnosis interval method estimated that 710 (3%) prevalent infections were attributable to blood or tissue transfer or blood factor treatment.

**DISCUSSION**

In this paper two indirect approaches to estimating the number of prevalent HIV infections in England and Wales at the end of 1993 are described. Although the methods use information from different sources they produce similar estimates and suggest that around 20000 to 23000 people were alive with HIV infection in England and Wales at the end of 1993 including those who were unaware of their infection status (Table 2). This compares with an estimated 22910 prevalent infections estimated using the direct method [9, 11]. About 57% of these infections were attributed to sexual intercourse between men, 9% to injecting drug use, 32% to sexual intercourse between men and women and 3% to blood factor treatment or...
blood and/or tissue transfer (Table 2). Using the same methods, estimates of prevalent HIV infections in England and Wales at the end of 1991 were about 1000 to 1500 higher [4] although there were almost 3000 deaths in the 2-year interval [9], suggesting there may have been around 1500–2000 incident HIV infections between 1991 and 1993.

The proportion of individuals who were unaware of their infection status is associated to a large extent with exposure category. In particular, undiagnosed HIV infections in those believed to have been exposed to infection through heterosexual intercourse were important, perhaps representing over half of all infections in persons exposed in this way. Such discrepancies in awareness of infection status by exposure group may be attributable to different perceptions of risk. If those who were exposed to infection through heterosexual intercourse are less likely to recognise themselves as being ‘at risk’, they may be less likely to seek a voluntary HIV test. The reasons for HIV testing in infected people were studied in 11 states and cities of the USA, and it was found that people infected with HIV through heterosexual contact were much less likely to have been tested because they thought themselves at risk, compared to those infected through sexual intercourse between men or injecting drug use [17].

The methods described involve assumptions and may contain biases which could have a significant influence on the estimates of infection prevalence. They are particularly sensitive to (1) biased information on HIV testing history and (2) reporting delay artefacts for HIV diagnoses, AIDS diagnoses and deaths in HIV-infected people.

Using the diagnosis interval methods, the date of first positive HIV test result may be inaccurate and in some cases the first known positive test may have been assumed to be at AIDS diagnosis. This would result in an underestimate of the number of HIV infections which were diagnosed prior to AIDS and would have the effect of overestimating the number of prevalent HIV infections.

Sampling bias in the unlinked anonymous surveys could seriously influence the estimate of prevalent HIV infections using the test history method. The assumption that IDUs with AIDS are unlikely to be attending drug centres may be invalid although information on this is limited. If substantial numbers of IDUs with AIDS are attending drug centres this could produce an overestimate of the proportion of infected individuals who have had their infection diagnosed, since everyone with an AIDS diagnosis will have had their HIV infection diagnosed. An overestimate of the proportion tested would tend to underestimate HIV infection prevalence.

A similar bias may occur within the data obtained from the GUM unlinked anonymous survey [1]. Infected individuals who attended GUM clinics are more likely to have been offered and to have received a voluntary and confidential HIV test than those who did not attend, which could result in underestimates of prevalent infections. This may explain why a much larger proportion of IDUs attending GUM clinics were aware of their infection status than was estimated using data from the IDU surveys and the ‘diagnosis interval’ method. HIV infection prevalence in homosexual men attending GUM clinics has been shown to be higher than in those who do not attend [18, 19] probably because the former group represents a more sexually active population who are acquiring sexually transmitted infections [18]. Similarly, HIV prevalence in heterosexuals (who were not known to have injected drugs) attending GUM clinics is considerably higher than in pregnant women at delivery [1]. Testing history in GUM attenders may not, therefore, represent that of the wider population of homosexual and heterosexual men and women, but it may be more representative of those who are infected with HIV.

To account for reporting artefacts, reports were adjusted using all available information on reporting patterns [9, 14, 15] (J. Mortimer, personal communication). As the methods are based on the number of prevalent reported HIV infections they are particularly sensitive to the accuracy of information on deaths in HIV-infected persons but this data is relatively robust.

Despite these reservations, the indirect methods described have particular merits. Unlike the direct method, they do not require data on the size of particular populations ‘at risk’ (such as injecting drug users) which is notoriously difficult to estimate. Indirect methods could be of immediate use to the many countries which operate national reporting of HIV diagnoses and the ‘diagnosis interval’ method would be particularly suitable for the majority of countries which do not have unlinked anonymous survey data. Finally, the indirect methods are cheap and simple applications of surveillance data which produce estimates of prevalent HIV infections which compare favourably with those produced by more complex methods.
Comment by authors at proof stage. Given the marked effects of highly active antiretroviral therapy becoming apparent in the industrialized world throughout 1997 and early 1998, it is unlikely that the assumption underlying the diagnosis interval method, that progression to AIDS is equal for individuals with long and short diagnosis intervals, holds beyond 1996 in these countries.

ACKNOWLEDGEMENTS

We are grateful to Professor G. V. Stimson for allowing us to use data from the survey of injecting drug users co-ordinated by the Centre for Research on Drugs and Health Behaviour. We also acknowledge the continued contribution of clinicians and microbiologists who report AIDS cases and HIV infections to the PHLS AIDS Centre on a voluntary and confidential basis and all those involved in the unlinked anonymous HIV prevalence monitoring programme.

REFERENCES


