How well do antenatal clinic (ANC) attendees represent the general population? A comparison of HIV prevalence from ANC sentinel surveillance sites with a population-based survey of women aged 15–49 in Cambodia

Vonthanak Saphonn, Leng Bun Hor, Sun Penh Ly, Samrith Chhuon, Tobi Saidel and Roger Detels

Background The purpose of this study was to evaluate whether HIV-1 prevalence among antenatal clinic (ANC) attendees in Cambodia provided a reasonable estimate of HIV-1 prevalence among all women 15–49 years.

Methods Antenatal clinic attendees in five HIV sentinel surveillance sites (five provinces) were selected by consecutive sampling (n = 1695). The population survey of females by household was carried out in the same five areas. Household females aged 15–49 years were selected using a three-stage cluster sampling design (n = 3066). Serum-based HIV ELISA testing was done for both ANC attendees and household females. The HIV prevalence for ANC attendees and household females were compared by age group and urban versus rural location.

Results The overall prevalence of HIV-1 infection among ANC attendees (1.62%, 95% CI: 1.26–1.98) was similar to the overall prevalence obtained from the general population of household females (1.24%, 95% CI: 0.92–1.55) in the same catchment areas in Cambodia. In the rural areas, the overall HIV prevalence among ANC attendees (2.18%, 95% CI: 1.59–2.77) was significantly higher than among the household females (0.86%, 95% CI: 0.49–1.23) after adjustment for age distribution and education level. In the 15–24 age group in rural areas, the HIV prevalence of ANC women was 2.71% (95% CI: 0.96–4.46) compared with 0.77% (95% CI: 0.02–1.53) in household females.

Conclusions Although ANC data can be used to estimate trends over time, it should be realized that ANC data may overestimate the actual prevalence in the younger age group in rural areas in Cambodia.

Keywords Cambodia, HIV, prevalence, sentinel surveillance, population survey

Accepted 3 October 2001
The first round of HIV sentinel surveillance started in 1994. The HIV sentinel surveillance system plays a major role in helping the Ministry of Health to monitor the epidemic trends of the disease and, more importantly, it provides information that helps to determine which groups of people are at risk, facilitating the development and prioritization of strategies for prevention and care programmes. The results from the surveillance programme have been used to estimate the trends of HIV infection in the sentinel groups themselves (or populations they are thought to represent) and to make national estimates and future projections that are useful for planning resources to combat the epidemic.

The HIV surveillance programme in Cambodia has expanded from only eight towns and provinces in 1994 to 20 towns and provinces in 1999, and from only urban populations to suburban populations and rural populations. Pregnant women attending antenatal clinics (ANC attendees) are one of the main sentinel surveillance groups. Data from ANC attendees (with data from other sources, such as blood donors and tuberculosis patients) have been used to estimate the prevalence in the general population. It was assumed that ANC attendees were representative of all adult women.

As improvements have been made in the sentinel surveillance system from year to year, the populations and the sampling methodologies have changed, making it difficult to analyse trends. Since less than one-third of pregnant women in Cambodia received antenatal care in government-supported clinics and delivered in a hospital, an important question is whether ANC attendees are indeed representative of the general population of women in Cambodia. In order to answer this question, a comparison of HIV prevalence was made between ANC attendees and prevalence data obtained from a population survey of women in the same catchment area. Although a population survey might provide a clearer picture of the epidemic in the general population, it is more time consuming and costlier than sentinel surveillance, and, thus, cannot be carried out routinely for providing trend data.

Several studies comparing sentinel surveillance data with population survey data confirmed that distortion of results is common. Some of the studies showed that antenatal data tends to give a lower estimate of HIV-1 prevalence in comparison to that from the general population females. Others found that ANC data tends to overestimate infection in teenagers aged 15–19 years, and underestimate the HIV infection in the 30 years or over age group. However, 9 years of experience from the Kagera region of Tanzania indicated that general population trend estimates can be generated using sentinel surveillance data based on pregnant women visiting an antenatal clinic for the first time during a given pregnancy. Several studies also concluded that antenatal data across the entire 15–49-year-old age range often provides remarkably robust estimates of HIV levels in the general population in mature epidemics. Recently another study suggested that for populations with low contraception usage, there is a procedure to adjust ANC data for improved estimates of HIV prevalence among women in sub-Saharan Africa.

In 1999, the National Center for HIV/AIDS, Dermatology and STD conducted a population survey of HIV prevalence in women, in addition to our regular sentinel surveillance programme in antenatal women in five provinces, to compare sentinel surveillance and population survey-derived data. The main goal of this article is to compare HIV-1 prevalence observed among ANC attendees from sentinel sites with a random sample of household females in the same catchment area during the same time period.

**Methods**

**Sampling design**

**Population survey**

The study was conducted in five provinces. A probability sample of 60 clusters of 10 randomly selected participants was selected from each province. Thirty clusters were selected from provincial towns considered to be urban areas. The same number of clusters was selected from the remaining districts or so-called rural areas. All household female members aged 15–49 years were eligible for the study.

A three-stage cluster sampling design was used. Geographical areas with fixed boundaries (villages) were used as the primary sampling units or ‘clusters’ in the initial stage of sample selection; then equally sized, smaller segments of 5–15 households (depending on the province), were selected by equal probability at the second stage. Finally, all the eligible respondents in the selected segments were asked to participate in the study. This sampling plan was designed to yield a self-weighted sample. A sampling frame of villages, with measures of size, was developed using the available 1993 census data, with updated information from the local administration office. Size was measured by the total population in the village.

A field test was conducted ahead of time to determine the number of eligible respondents that could be expected from each segment. Since a take-all approach was used for each segment, the sample size was inflated ahead of time to account for the estimated per cent of: (1) refusals (those who were contacted but refused to participate in the study), (2) non-residing residents (those who were away from home for an extended period of time), and (3) residents who were not at home (those who were residing at home but who were still unavailable after three separate call-backs at different times). This inflation was necessary to ensure that the required number of respondents would eventually be included in the sample while still maintaining the self-weighted design. During the actual survey, for residents who were not at home, three call-backs at different times were carried out in order to minimize the bias that could result by skipping residents who happened not to be at home when the survey team arrived. Participants were informed that the results of the HIV test would be used strictly for research purposes, that they would not receive the results of the test, and that data would be handled anonymously. They were also informed about their right to refuse to participate in this study.

**Antenatal clinic attendees**

Pregnant women attending a designated public ANC for their first visit for any pregnancy were eligible for the study. Three hundred pregnant women per province were selected consecutively from the sentinel sites. Voluntary anonymous testing was done at the ANC.

**Sample size calculation**

In both groups, we assumed from prior surveys that the best estimate of the true proportion infected in the sampled
population was 3%. For the population survey, the sample size was determined using the CSURVEY module in the Epinfo software package (Centers for Disease Control and Prevention, Atlanta, Georgia, USA) to make sure that the sample size of 600 would be adequate to assess the prevalence of HIV infection in each study site, within plus or minus 2%. The design effect was assumed to be 1.5 for the purpose of sample size determination. For sentinel surveillance, the sample size was calculated to be able to derive a 95% CI with a margin of error no greater than 2%.

HIV-1 serology

The WHO recommended testing strategy type II was followed.\(^1\)\(^7\) All samples that were positive at the first test (Particle agglutination test: Serodia 1–2\(^\circ\)) were then re-tested with the ELISA test (Genelavia Mixt\(^\circ\)). Individuals positive for both tests were considered to be HIV-infected. Discordant sera were considered negative. Testing in the central laboratory in Sihanouk Hospital, Phnom Penh, started about one month after the start of the survey. During the process, quality control of the laboratory was strictly maintained in collaboration with the Pasteur Institute in Cambodia. The Pasteur Institute sent out quality control samples to the laboratory and testing of the samples was done at the same time as the other specimens. At the provincial level sera were stored in a –20°C freezer. Coolers were used to transport the sera from the provinces to the central laboratory.

Validation strategy and statistical analysis

Data from the ANC attendees from sentinel surveillance sites were compared with data from the household female population survey. The HIV prevalences for ANC attendees and household females were compared by age group and urban versus rural location. Epinfo was used for analysis. Since there were some flaws in the data collection process, our population survey data did not have complete information about the cluster to which each individual belonged. Therefore, it was not possible to perform a proper cluster analysis and it was necessary to assume that the population survey data were drawn based on simple random sampling. However, a sensitivity analysis was done using design effects of 1.5, 2 and 3. This design effect was assumed when calculating the \(P\)-value for the Z test of two proportions, which was used to calculate the difference in prevalence between the general population of women and the ANC attendees.

Results

A total sample of 1695 ANC attendees and 3066 household females in the community were included. The overall refusal rate was 5% and 7% for ANC attendees in the urban areas and the rural areas, respectively. The refusal rate among the sample of household females was higher, ranging from 10% in the rural areas to 20% in the urban areas.

Table 1 shows the characteristics of the ANC attendees and the household females in our population survey. Means were used as a measure of central tendency. The ANC attendees were significantly younger (26.74 years versus 29.9 years, \(P < 0.001\)) and had more education than the household females in the population survey (6.45 years of schooling versus 4.34 years, \(P < 0.001\)). After adjustment for the five provinces, the overall HIV prevalence in household females was lower than that in ANC attendees (1.24% versus 1.62%). This difference was not statistically significant. However, in rural areas the overall HIV prevalence in household females was significantly lower that in ANC attendees (0.92% versus 2.32%, \(P = 0.015\)). In Table 2, after adjusting for age group, the difference in HIV prevalence between the two groups was significant (\(P = 0.01\)). When adjusting for age group and education level the difference in HIV prevalence between the two groups was significant too (\(P = 0.02\)). This difference remained statistically significant after we took design effects of 1.5 and 2 into consideration. Age group specific comparisons showed that the younger group (15–34 years) of ANC attendees had a significantly higher HIV prevalence than the household females in the population survey (\(P = 0.05\)). After taking a design effect of 1.5 into account, the \(P\)-value was 0.08. In comparisons by education level, we found that among the women who had > 5 years of schooling (more than primary school level), ANC attendees had significantly higher HIV prevalence (2.35%, 95% CI: 1.69–3.02) than

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Characteristics of antenatal clinic (ANC) attendees and household females</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristics</strong></td>
<td><strong>ANC attendees (N = 1695)</strong></td>
</tr>
<tr>
<td><strong>Province</strong></td>
<td><strong>No. tested</strong></td>
</tr>
<tr>
<td>Battambang</td>
<td>374</td>
</tr>
<tr>
<td>Kampong Cham</td>
<td>566</td>
</tr>
<tr>
<td>Kamot</td>
<td>255</td>
</tr>
<tr>
<td>Kratte</td>
<td>200</td>
</tr>
<tr>
<td>Takeo</td>
<td>300</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td><strong>Mean years of schooling (SD)</strong></td>
</tr>
<tr>
<td><strong>Urban areas</strong></td>
<td>878</td>
</tr>
<tr>
<td><strong>Rural areas</strong></td>
<td>817</td>
</tr>
<tr>
<td><strong>Mean years of schooling (SD)</strong></td>
<td><strong>2.64 (4.58)</strong></td>
</tr>
<tr>
<td><strong>Mean age (SD)</strong></td>
<td><strong>26.74 (6.22)</strong></td>
</tr>
<tr>
<td><strong>Total sample</strong></td>
<td>1695</td>
</tr>
</tbody>
</table>
| \(^a\) Overall HIV prevalence in two groups after direct adjustment for size of province.
This difference persisted with a design effect as high as 3.

Table 3 compares HIV prevalence in the household females with that of ANC attendees in the urban areas. There was no difference between the two groups for every age group and education level.

Table 4 compares HIV prevalence in the household females with that of ANC attendees in the rural areas. After adjusting for age, overall HIV prevalence in the household females in the rural areas was significantly lower than in ANC attendees in the rural areas (2.23% versus 0.88%, P < 0.003). The difference remains statistically significant even though we assume a design effect of as high as 3. After adjusting for age and education level, overall HIV prevalence in the household females in the rural areas was still significantly lower than in ANC attendees in the rural areas (2.18% versus 0.86%, P < 0.003). The difference remains statistically significant even when we assume design effects of 1.5, 2 and 3. When stratifying on age group, we found
that the difference was greatest in the younger age group, 15–24 years (0.77% versus 2.71%, \( P = 0.04 \)). With a design effect of 1.5 the \( P \)-value is 0.05. Among those with >5 years of education, we found that the ANC attendees had significantly higher HIV prevalence than household females (5.29% versus 3.93%). This difference persisted with a design effect as high as 3. The difference was most important among women aged 15–24 years.

When looking at the urban and rural differential, adjusting for age and education level, the difference was significant in ANC attendees (1.34% versus 2.54%, \( P = 0.01 \)). The difference was also significant in household females (1.52% versus 0.90%, \( P = 0.02 \)). This difference persisted with a design effect of 1.5 and became borderline significant with a design effect of 2.

**Discussion**

Although the overall prevalence of HIV-1 infection among ANC attendees was similar to the prevalence obtained from the sample of the general population of household females in the same catchment areas in Cambodia, when the comparison was stratified by age group and area (urban versus rural), the HIV prevalence obtained from ANC attendees was significantly higher than that obtained from the general population of household females among the younger age population (15–24) in rural areas. The HIV prevalence in ANC attendees age 15–24 was significantly higher than among household females for those with >5 years of schooling. This significant difference persisted even when we took into consideration design effects as high as 3 in the younger group of household females. Women attending ANC are, by definition, sexually active, whereas not all women in the general population, especially in the youngest age group, are sexually active. Thus, it is not unexpected that HIV prevalence is higher among ANC attendees.

Data from ANC attendees are probably not generalizable because, according to a national health survey in 1998, only one-third of Cambodian women who gave birth over the past 5 years received antenatal care from a medically trained person. The percentages are higher for urban women, women who are literate/educated, and women with higher socioeconomic status. Among women who gave birth living in remote or isolated rural areas, and among women who were poor and/or illiterate, less than one-fifth received antenatal care from a trained provider.11

In our study, we attempted to use narrower age groups (15–19, 20–24, 25–29, >30 years) but the cell sizes, when looking at urban versus rural differences after controlling for age group and education level, became too small to permit meaningful comparisons. Therefore, we decided to select the age groups 15–24, 25–34, >35 years. We did not find any significant difference between the ANC data and household female data in the >35 age group. The possible reason might be due to the small sample size of this age group which may not allow the study to have enough power to detect a difference. Similar studies employing larger sample sizes are recommended to assess the difference of HIV infection in younger (15–19) and older age groups (>35).

On the other hand, the lower response rate in household females should also be considered when comparing the data from these two groups. Since we did three call-backs, the women whom we missed in the study were more likely to work outside of the village (far from home) and possibly were away from home for a period of time working in the city. Most studies have shown that migrant women are more likely to be at higher risk than those who stay with their family in the villages. In addition, in the Cambodian culture, the concepts of confidentiality

### Table 4 Age group and location comparison of antenatal clinic (ANC) attendees data and household females data in rural area

<table>
<thead>
<tr>
<th>Education level (0–5 years)</th>
<th>ANC attendees</th>
<th></th>
<th>Household females</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. tested</td>
<td>No. positive</td>
<td>Prevalence (95% CI)</td>
<td>No. tested</td>
<td>No. positive</td>
</tr>
<tr>
<td>15–24</td>
<td>158</td>
<td>4</td>
<td>2.53 (0.08–4.98)</td>
<td>320</td>
</tr>
<tr>
<td>25–34</td>
<td>202</td>
<td>2</td>
<td>0.99 (0.00–2.36)</td>
<td>303</td>
</tr>
<tr>
<td>≥35</td>
<td>92</td>
<td>1</td>
<td>1.09 (0.00–3.21)</td>
<td>394</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.52 (0.90–2.15)(^a)</td>
<td></td>
</tr>
<tr>
<td>Education level (&gt;5 years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15–24</td>
<td>174</td>
<td>5</td>
<td>2.87 (0.39–5.36)</td>
<td>198</td>
</tr>
<tr>
<td>25–34</td>
<td>157</td>
<td>6</td>
<td>3.82 (0.82–6.82)</td>
<td>218</td>
</tr>
<tr>
<td>≥35</td>
<td>34</td>
<td>1</td>
<td>2.94 (0.00–8.62)</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.29 (2.11–4.47)(^d)</td>
<td></td>
</tr>
<tr>
<td>Total (adjusted)</td>
<td></td>
<td></td>
<td>2.23 (1.64–2.83)(^g)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.18 (1.59–2.77)(^i)</td>
<td></td>
</tr>
</tbody>
</table>

Note: all the values in the table are calculated using a simple random sampling assumption.

\( a \) Direct adjustment for age distribution.

\( b \) Direct adjustment for age distribution and education level.

\( c \) With design effect of 1.5, \( P = 0.02 \); with design effect of 2, \( P = 0.02 \); with design effect of 3, \( P = 0.02 \).

\( d \) With design effect of 1.5, \( P < 0.003 \); with design effect of 2, \( P < 0.003 \); with design effect of 3, \( P = 0.007 \).

\( e \) With design effect of 1.5, \( P < 0.003 \); with design effect of 2, \( P < 0.003 \); with design effect of 3, \( P = 0.003 \).

\( f \) With design effect of 1.5, \( P < 0.003 \); with design effect of 2, \( P = 0.005 \); with design effect of 3, \( P = 0.01 \).
and privacy are not clearly understood by the public. Even though we explained that the testing would be done anonymously, there were still some people who suspected themselves to be infected and, thus, avoided participation in the study because they were afraid of having their status disclosed. If the non-responders among the sample of household females were at higher risk, the HIV prevalence data for household females in this study might underestimate the true prevalence for the population.

In the sensitivity analysis, the range of design effects that we incorporated was based on the design effects observed from the behavioural surveillance survey in Cambodia in the year 2000. For the individual questions, the design effect varied from 0.9 to 3. The answer to questions such as, ‘Have you paid to have sex in the last year?’ yielded a lot of variation within and between clusters, which led to an elevated design effect of 3. However, we think that for HIV prevalence the variation between and within the cluster is not that high. We prefer the assumption of a design effect of 1.5, but we have used a range for design effect in the sensitivity analysis.

From previous studies, it has been shown that HIV infection reduces fertility. Thus, some infected women in the reproductive age range would be in the general population but not the antenatal population. Thus, the HIV prevalence from ANC attendees might be an underestimate of the infection in the general female population.

### Conclusion

Because it is not feasible to repeat household surveys annually, ANC data can be used to estimate trends over time assuming that the margin of error remains constant over time. It should be realized, however, that ANC data may overestimate the actual prevalence in the younger age group. The quality of data from pregnant women attending ANC could provide more representative information if some improvement to the system could be made, such as reducing selection bias by encouraging and motivating as many pregnant women as possible to attend ANC, and by including private ANC in the sentinel surveillance sample.

### Acknowledgements

This study was partly funded by Fogarty International Center grant TW00013. I would like to thank the National Center for HIV/AIDS, Dermatology and STD (NCHADS) for the use of their HIV Sentinel Surveillance Program 1999 data. The authors also wish to acknowledge Professor William G Cumberland, Chair, Department of Biostatistics, University of California, Los Angeles School of Public Health, for his comments and advice on the data analysis. In addition, the authors would like to thank Dr Kaoeun Chettra, Dr Lim Yi, Dr Theng Thithara, and all the provincial AIDS programme staff members for their endless effort doing the data collection and making this survey possible.

### References

8. Saidel T. Methodological issues in the Cambodia HIV sentinel surveillance and the representativeness of antenatal clinic women. Presentation at the Monitoring the AIDS Pandemic (MAP) Network...


