The total number of estimates of the relative risk associated with exposure to risk factor 1 that met the stated objectives of precision in estimation when the population sero-prevalence rate was varied is presented in Table 4, for responders only: a probability of nonresponse of 20% was applied to infected adults and a rate of only 5% used to determine the response status of noninfected adults. The number of EPI-like estimates falling within the range defined by the true rate ±0.10 tended to be quite similar to those obtained using SRS. With either method of sampling, the number of estimates meeting this objective rose as the level of infection increased.

The total number of estimates within the range defined by the true relative risk ±20% times the actual value, indicated that either sampling strategy was quite capable of meeting this objective with a high degree of confidence at virtually every sero-prevalence level evaluated. Even in the lower range of infection, which is more likely for HIV infection, the EPI-like method performed nearly equally to or better than SRS.

Varying the level of infection appeared to have little impact on construction of confidence intervals. Confidence interval estimates generated using either method of sampling included the true population value in over 90% of samples. Overall, the data observed in this table fail to demonstrate a clear advantage of one sampling technique over the other in meeting the stated objectives.

The total number of estimates of the relative risk of exposure to risk factor 2 that met stated objectives of precision in estimation when the population sero-prevalence rate was varied is presented in Table 5. It should be recalled that the magnitude of association between exposure and disease status for the second risk factor was set at approximately 3.0.

### Table 4. Number of Estimates of the Population Relative Risk of Exposure to Risk Factor 1, Based Solely on Responding Adults, That Met Stated Objectives in Precision When True Sero-prevalence Rate Was Varied

<table>
<thead>
<tr>
<th>Sero-prevalence rate (Taux de séropérvance (%))</th>
<th>RR ± 0.1</th>
<th>RR ± 20% (RR)</th>
<th>RR within CIE (RR dans un intervalle de confiance donné)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EPI PEV</td>
<td>SRS SAS</td>
<td>EPI PEV</td>
</tr>
<tr>
<td>2</td>
<td>240</td>
<td>222</td>
<td>383</td>
</tr>
<tr>
<td>4</td>
<td>313</td>
<td>318</td>
<td>456</td>
</tr>
<tr>
<td>6</td>
<td>394</td>
<td>365</td>
<td>488</td>
</tr>
<tr>
<td>8</td>
<td>417</td>
<td>415</td>
<td>498</td>
</tr>
<tr>
<td>10</td>
<td>416</td>
<td>436</td>
<td>492</td>
</tr>
<tr>
<td>15</td>
<td>459</td>
<td>467</td>
<td>500</td>
</tr>
<tr>
<td>20</td>
<td>486</td>
<td>487</td>
<td>500</td>
</tr>
<tr>
<td>25</td>
<td>487</td>
<td>491</td>
<td>500</td>
</tr>
<tr>
<td>30</td>
<td>495</td>
<td>497</td>
<td>500</td>
</tr>
<tr>
<td>40</td>
<td>499</td>
<td>500</td>
<td>500</td>
</tr>
</tbody>
</table>

* Expanded Programme on Immunization – Programme élargi de vaccination.

* Simple random sampling – Sondage aléatoire simple.

### Table 5. Number of Estimates of the Population Relative Risk of Exposure to Risk Factor 2 That Met Stated Objectives in Precision When True Sero-prevalence Rate Was Varied

<table>
<thead>
<tr>
<th>Sero-prevalence rate (Taux de séropérvance (%))</th>
<th>RR ± 0.1</th>
<th>RR ± 20% (RR)</th>
<th>RR within CIE (RR dans un intervalle de confiance donné)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EPI PEV</td>
<td>SRS SAS</td>
<td>EPI PEV</td>
</tr>
<tr>
<td>2</td>
<td>113</td>
<td>95</td>
<td>462</td>
</tr>
<tr>
<td>4</td>
<td>132</td>
<td>163</td>
<td>471</td>
</tr>
<tr>
<td>6</td>
<td>174</td>
<td>179</td>
<td>497</td>
</tr>
<tr>
<td>8</td>
<td>183</td>
<td>175</td>
<td>483</td>
</tr>
<tr>
<td>10</td>
<td>196</td>
<td>221</td>
<td>492</td>
</tr>
<tr>
<td>15</td>
<td>199</td>
<td>265</td>
<td>499</td>
</tr>
<tr>
<td>20</td>
<td>179</td>
<td>237</td>
<td>499</td>
</tr>
<tr>
<td>25</td>
<td>173</td>
<td>322</td>
<td>497</td>
</tr>
<tr>
<td>30</td>
<td>171</td>
<td>381</td>
<td>496</td>
</tr>
<tr>
<td>40</td>
<td>129</td>
<td>403</td>
<td>486</td>
</tr>
<tr>
<td>50</td>
<td>107</td>
<td>404</td>
<td>481</td>
</tr>
</tbody>
</table>

* Expanded Programme on Immunization – Programme élargi de vaccination.

* Simple random sampling – Sondage aléatoire simple.

The number of estimates of the relative risk that fell within the range defined by the true rate ±0.10 was generally rather small using either method of sampling. However, if the range of interest was instead defined on the basis of the true value ±20% times the actual value, the total number of estimates meeting this objective in precision was generally shown to increase along with increases in the seroprevalence rate. Indeed, either sampling technique appeared quite capable of meeting this objective in precision with a high degree of confidence at the levels evaluated.

The ability of either method of sampling to construct confidence intervals which included the true relative risk tended to diminish as the prevalence of infection rose; this was particularly true of estimates employing the EPI survey strategy. When the infection rate exceeded 15%, the EPI-like method generated confidence intervals which included the actual parameter value for less than 70% of the samples.

The total number of estimates of the relative risk of exposure to the second risk factor that met the stated objectives of precision in estimation when the population seroprevalence rate was varied is reported in Table 6, for responders only.

The total number of estimates within the range of the actual value ±0.10 indicated that neither method of sampling was capable of meeting this objective. The number of estimates within the range defined by the true relative risk ±20% times the actual value demonstrated marked improvement over the previous study objective; the data suggest that either sampling strategy was capable of meeting this objective with a high degree of confidence.

Varying the level of infection appeared to substantially influence the construction of confidence interval estimates. Generally, either method of sampling enabled confidence intervals that included the true population value with a high degree of confidence only when the infection rate did not exceed 20%. However, throughout the range of infection examined, the observed data generally failed to demonstrate a clear advantage of one sampling technique over the other with regard to confidence interval estimation.

Discussion

The use of the EPI-like survey technique for estimating relative risks was evaluated with respect to three different objectives in precision in estimation. The levels of precision employed for the first and second objectives included estimation of the relative risk (RR) within the range defined by the true level ±20% times the actual RR and estimation of the RR to within the narrower range defined by the RR ±0.10. A 95% confidence interval was used in assessing the third objective.

In the absence of pocketing of infection, results obtained from five simulated populations failed to provide sufficient evidence to suggest that the EPI-like survey strategy, when applied at the second stage of sampling, was less able to meet the stated objectives in precision than SRS with respect to estimation of the population relative risk.

When pocketing of infection was included within the population, although differences were noted between the two methods with respect to the stated objectives in precision in estimation of relative risks, their magnitude was insufficient to recommend one method over the other. Moreover, pocketing of infection did not appear to differentially affect the precision of the two survey methods.

Though not included in the tables, when sampling from the various study populations, the number of adults selected per household for inclusion in the survey (1 vs. all) did not appreciably affect the precision in estimation of relative risks using either method of sampling.

The observed estimates of relative risks based solely on responding adults indicated that while the total number of estimates meeting the stated objectives tended to decrease, the precision of the two sam-

### Table 6: Number of Estimates of the Population Relative Risk Associated with Exposure to Risk Factor 2, Based Solely on Responding Adults, That Met Stated Objectives in Precision When the True Seroprevalence Rate Was Varied

<table>
<thead>
<tr>
<th>Seroprevalence rate (%)</th>
<th>RR ± 0.1</th>
<th>RR ± 20% (RR)</th>
<th>RR within CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EPI* - PEV*</td>
<td>SRS* - SAS*</td>
<td>EPI* - PEV*</td>
</tr>
<tr>
<td>2</td>
<td>103</td>
<td>90</td>
<td>449</td>
</tr>
<tr>
<td>6</td>
<td>98</td>
<td>121</td>
<td>451</td>
</tr>
<tr>
<td>8</td>
<td>156</td>
<td>152</td>
<td>494</td>
</tr>
<tr>
<td>10</td>
<td>179</td>
<td>181</td>
<td>491</td>
</tr>
<tr>
<td>15</td>
<td>182</td>
<td>188</td>
<td>496</td>
</tr>
<tr>
<td>20</td>
<td>174</td>
<td>169</td>
<td>493</td>
</tr>
<tr>
<td>25</td>
<td>191</td>
<td>211</td>
<td>495</td>
</tr>
<tr>
<td>30</td>
<td>123</td>
<td>126</td>
<td>488</td>
</tr>
<tr>
<td>40</td>
<td>144</td>
<td>76</td>
<td>485</td>
</tr>
<tr>
<td>50</td>
<td>103</td>
<td>9</td>
<td>433</td>
</tr>
</tbody>
</table>

* Expanded Programme on Immunization – Programme élargi de vaccination.
* Simple random sampling – Sondage aléatoire simple.
pling methods generally did not appear to differ markedly. Differences in precision in estimation of relative risks using the two sampling methods were of little practical significance when nonresponse was differentially or non-differently distributed, even in the presence of pocketing of infection.

In order to investigate the impact of other forms of potential bias on precision in estimation when using the EPI-like sampling design, sampling was also conducted using populations constructed with varying levels of misclassification of exposure status. For exploratory purposes, a range of values was chosen for designating misclassification of exposure. Differences in precision between the two methods when estimating the relative risk where there was misclassification of exposure were found to be of little practical significance.

To determine the effect on precision in estimation that results from varying the seroprevalence rate, the simulation was carried out using a standard population constructed with different levels of infection; infection rates ranging between 2% and 50% were investigated. Varying the level of infection did not appear to differentially affect the two means of survey sampling with regard to estimation of the population relative risk when it was set equal to 1.0. In this case, the difference in the number of estimates meeting each of the stated objectives using the two survey methods did not differ widely enough to be of any practical significance.

In attempting to estimate a larger magnitude of risk (RR=4.0), the SRS strategy was generally better able to meet the objectives in precision, especially with higher seroprevalence rates.

Obviously, the real world is much more complicated than can be represented in a computer simulation. While the resulting simplifications should be viewed as an arbitrary representation of reality, they can potentially influence the interpretation of the results. For instance, there was no attempt to construct clusters with varying numbers of households and the chosen cluster size of 600 households is, no doubt, larger than many areas where such surveys may be conducted.

Pocketing of infected adults was also represented in an exaggerated fashion. Pockets of infection were produced in the simulation models reported by infecting 75% of all adults residing in households within the pockets and none elsewhere within the cluster, although the program allowed other choices. Such extremes are likely to be rare in real populations.

The computer algorithm used to simulate movement of interviewers from house to house represents yet another example of the simplification of reality. This algorithm was based on an orderly arrangement of households within a matrix which clearly does not reflect conditions typically encountered in the field. In actual field operations, non-sampling errors (such as interviewer, recall, and reporting bias) can be a significant problem. The simulation attempted to evaluate the impact of nonresponse and misclassification bias only. It was assumed that the potential for a selection bias was minimized through repeated sampling from the populations created. Aside from these potential biases, no attempt was made to address other potential sources of non-sampling errors that may arise under field conditions.

While attempts were made to reflect a number of population characteristics in the simulation, of necessity it was not possible to include all features that characterize real populations. Moreover, since each execution of the simulation required considerable time to complete, only certain combinations of these characteristics were explored. Nevertheless, for the characteristics that were included, a range of possible values was examined.

Having stated the various limitations of the simulation, it is important to note that it is unlikely that any of these would differentially affect results obtained from the two methods of sampling.

In summary, if the principal outcome of interest in a particular study is to estimate the relative risk associated with exposure to some factor with a high degree of precision, the EPI-like survey method appears to be a reasonable alternative to the use of SRS at the second stage of sampling. In virtually all of the situations evaluated, estimates of the relative risk obtained using the two sampling methods failed to demonstrate a clear advantage in the use of one survey design over the other. However, the added cost and difficulty encountered in implementing SRS at the second stage of sampling certainly warrant further consideration of the EPI methodology for use in estimating relative risks.

Evidence from this study also suggests that sampling involving selection of only one adult per household may improve the precision of estimates derived using the EPI-like method. However, since selection of only one adult per household would require that a greater number of households be visited, any gain in precision must be weighed against the likely increase in cost of household visitation.

SUMMARY

Precision in estimation of relative risks using a standardized sampling method proposed by the WHO Global Programme on AIDS was evaluated using a Monte Carlo model simulating actual populations; the proposed survey design represents a modification of the methodology used by the WHO Expanded Programme on Immunization (EPI) to estimate immunization coverage among children. This study suggests that in actual populations the proposed survey strategy is a reasonable alternative to the use of simple random sampling (SRS) at the second stage of cluster sampling. Although varying such population characteristics as the seroprevalence rate, nonresponse rate, and rate of misclassification of exposure failed to demonstrate a clear advantage of one method over the other, the added cost and difficulty of implementing SRS under field conditions warrant further consideration of the EPI-like methodology for use in estimating relative risks.
RÉSUMÉ
Evaluation de la méthodologie d’enquête du PEV
concernant l’estimation du risque relatif

On a évalué, au moyen d’un modèle de Monte Carlo
simulant des populations réelles, la précision de
l’estimation du risque relatif obtenue par la méthode
de sondage normalisée que propose d’utiliser le
Programme mondial OMS de lutte contre le SIDA. Le
plan d’enquête proposé est une variante de la
méthodologie utilisée par le Programme élargi de
vaccination (PEV) de l’OMS pour estimer la couver-
ture vaccinale des enfants. Il semble, au vu des
résultats, que dans le cas de populations réelles, la
méthodologie proposée puisse, au second degré
d’un sondage en grappes, se substituer valablement
au sondage aléatoire simple (SAS). Même si, en
faisant varier des caractéristiques de la population
telles que le taux de séroprévalence, le taux de
non-réponse et le taux d’erreur sur le type
d’exposition, on ne peut pas trancher en faveur de
l’une des deux méthodes, le coût supérieur et la
difficulté d’exécution du SAS sur le terrain justifient
que l’on étudie plus avant la méthodologie utilisée
par le PEV pour l’estimation du risque relatif.

REFERENCES — RÉFÉRENCES

1. HENDERSON, R. H. & SUNDARESAN, T. Cluster sam-
ppling to assess immunization coverage: a review
of experience with a simplified sampling method.
Bulletin of the World Health Organization, 60 (2):
253-260 (1982).

HENDERSON, R. H. & SUNDARESAN, T. Sondage par
grappe pour évaluer la couverture vaccinale:
examen d’une expérience sur une méthode
simplifiée de sondage (résumé). Bulletin de
l’Organisation mondiale de la Santé, 60 (2): 259-
260 (1982).

2. LEMESHOW, S. & ROBINSON, D. Surveys to measure
programme coverage and impact: a review of the
methodology used by the Expanded Programme
on Immunization. World health statistics quar-

LEMESHOW, S. & ROBINSON, D. Enquêtes sur la
couverture et l’impact des programmes: métho-
des quantitatives utilisées par le Programme
élargi de vaccination. Rapport trimestriel
de statistiques sanitaires mondiales, 38 (1): 65-75
(1985).

3. LEMESHOW, S. ET AL. A computer simulation of the
EPI survey strategy. International journal of epi-