The Numbers Game: The story behind the estimates

By Dr L.M. Nath

An annual feature of the HIV scene in India is the release of HIV estimates, both by the National AIDS Control Organisation (NACO) and the UNAIDS separately. These releases are usually followed by comments on their accuracy and relevance.

In 2007, the Union minister for health and family welfare, Anbumani Ramadoss, announced that new estimates found only 2.5 million people were living with HIV in India. This almost halved the previous year's estimates of 5.2 million infections. NACO explained the revised estimates had been arrived at by using a "better" method of calculating HIV numbers.

The implications of these statements were not fully understood by many. Especially those not technically up to date on the methodology of HIV estimation and HIV surveillance.
Several questions remained unanswered. For instance, people said they did not know, figures apart, whether HIV prevalence in the country actually increased or decreased. At worst, many said that NACO had engaged in juggling figures because of the criticism of the "inflated" figures to meet national or international pressures.

Is the reported difference really an outcome of a new methodology? If that is so, is the total quantum of infection in India static or is it increasing or decreasing? And if there is a real change, how much is it? Is it confined to particular groups or geographical areas or is it a generalised phenomenon? If the current year's data were to be analysed using the same methods as were used during the previous year, would the figure obtained be the same as last year, or would it be less or more?

The questions can be addressed only by recalculating the previous years' figures using the current methodology. In other words two sets of data can only be compared if the basic methodology remains the same for the two corresponding periods.

When the rates for the two years were calculated using the same methodology, the figure for the current estimate dropped to 4.84 million from the earlier 5.216. This is important and a welcome indicator that the epidemic is perhaps beginning to level off. Looking at the details of the prevalence detected at different 'sentinel' sites it becomes clear that the leveling has occurred in the south of India and is more pronounced in the higher risk categories.

It is also worth considering what the 'new methodology' was and why it produced such a drastically different estimate of the total infected persons in the country. Before we can do that some basics need to be explained:

India, like most countries in the world, does not count the number of infected persons in the country by testing everyone. The total number is estimated by using a mathematical process. The entire population is divided into several risk categories. The HIV prevalence in a small sample of each category is actually measured at what is known as a sentinel site and then extrapolated to give a probable figure for the entire population that falls in that particular category.

The risk categories selected are the general population, who are usually not at a serious risk of getting a sexually transmitted disease, including HIV. Additional groups, whose lifestyle may predispose them to a higher risk of HIV are included separately. In India the groups that are considered to have a higher risk

category include the following: injection drug users, truckers, men who have sex with men, sexually transmitted infection, clinic attendees and sex workers

As mentioned earlier, HIV prevalence in those persons visiting a "sentinel site" during a specified period is measured. These sentinel sites are identified at certain places across India and all those who visit these sites during September, October and November constitute the sample. NACO has specified that each site should include 300 persons in low risk categories and 200 in the high risk groups.

HIV prevalence is not distributed uniformly in India or indeed in any large country. To get a good idea of the HIV epidemic in the country, it is necessary therefore to establish many sentinel sites for each category and ensure that these represent as far as possible all geographic and risk zones in the country.

The crucial issue is what group should be tested to represent the "general population". This is important as this is by far the largest segment of the population. Because of the social implications of the disease and the resultant stigma and discrimination, it has generally not been considered ethical to identify persons who are HIV positive. Also, if people learnt that they would be tested for HIV, the group agreeing to the test may not be the ideal representative sample.

We would call this a biased sample or non-representative sample because those that volunteer may not be the same as those who do not agree to be tested for HIV. The solution for this dilemma used by epidemiologists is to do what is called 'anonymous unlinked testing'. This is when a small part of the blood, drawn for a legitimate medical reason is taken aside and given a code number that cannot be traced back to the individual. When this coded sample is tested, it is not a person being tested but just a blood sample. The purpose is not to find out who is HIV-positive but, rather, what proportion of the blood samples is HIV positive. This proportion - or prevalence - is then applied to the number of persons in that risk category.

The problem is that it is not easy to find a group that represents the general population who actually are subjected to blood tests for legitimate medical reasons. The solution used until recently was to make the assumption that pregnant women who go to antenatal clinics represent the general population.

This, of course, was a far from an ideal solution. There are several important differences between those that attend antenatal clinics and the general population of women. One, the age distribution is not typical of the population, pregnant women are younger. Secondly, by definition, pregnant women are both sexually active and not using condoms. For all these reasons we could expect a slightly higher prevalence of HIV in pregnant women. However, this was accepted as we could not think of a substitute that was more representative of the general population.

The figures that the Union health minister quoted came from using the additional data provided by the National Family Health Survey III, a nationwide comprehensive study of health indicators conducted every five years. For the first time this year, the NFHS added a blood test for HIV to its normal pattern of questions. The survey tested over 1.12 lakh blood samples from all over the country. So, for the first time we had data on HIV prevalence that were a truer representation of the general population.

This data could be used to get a better estimate of the general population HIV prevalence than we got from testing pregnant women alone. This was the basic difference between the method previously used and the methodology used last year. It is worth noting that whenever the estimates have been calculated by this 'new' methodology internationally, there have been drastic reductions in the estimate for that country.

Another change that occurred in 2007 was that the number of sentinel sites was increased sharply, from 700 odd to 1122 active sites. This was done to include every district in the country in sentinel surveillance. In the past, the vast majority of sites were in the high prevalence states and the large states in the north had fewer sites. Adding additional sites in the low prevalence areas may also have influenced the total estimates for India.

Finally, it is time to reconsider the true value of sentinel sites. Sentinel surveillance was not designed to help make estimates. It is unfortunate that the true purpose has generally been forgotten because of the interest in total HIV estimates. By comparing the prevalence in the same site with earlier figures, very good site-specific information can be obtained about HIV prevalence in that particular area. This data can and should be used for evaluation and planning purposes. It is unfortunate that not too many people look at data in that manner.

(Dr L.M. Nath is an epidemiologist and a former director of the All India Institute of Medical Sciences)