
Estimation of missing values

Sir--*The Lancet* is to be commended for publishing the [viewpoint by the AVANTI Steering Committee](#) .¹ As this group eloquently points out, the exclusion of missing data from analyses labelled intent-to-treat is misleading; therefore, data for all individuals must be included in such analyses, and conservative rules must be adopted to replace missing data values with assigned data.

We have wrestled with this issue for some time while deciding how to assign missing data values for survival analyses and for glutathione measurements in blood and T cells in an intent-to-treat HIV study that we recently completed. The solutions we evolved, which have been approved by our local statisticians, are more complete and more conservative than those proposed by AVANTI. Therefore, we suggest the following.

First, to estimate missing categorical responses, drug-group participants should be assigned the response corresponding to failure for the drug, whereas those in the placebo group are assigned the response for success. Second, to estimate missing survival data, placebo-group participants should be treated as having survived for the length of the study but those in the drug group should be treated as having died, and their time-to-death assigned as the median time-to-death for individuals in the drug group whose time-to-death is known. To avoid repeatedly assigning the same (median) value to a series of missing values, the sampling method described below could be adapted for assigning time-to-death here.

Third, to estimate missing values for quantitative responses, participants should be assigned values by sampling from the distribution of all known values--ie, drug and placebo groups combined. For example, in our study, we replaced missing baseline values with values selected randomly from the distribution of all known baseline values (drug group and placebo group combined). Similarly, we replaced the missing endpoint values by randomly sampling from the distribution of all known endpoint values. To reduce sampling error, we constructed 50 such datasets, analysed each independently for treatment effect, and combined the results by computing the median of the significance values obtained individually for the 50 sets.

**Leonore A Herxenberg, Wayne A Moore, Stephen C De Rosa*

*Genetics Department, Stanford University School of Medicine, Stanford, CA 94305, USA(e-mail:
LeeHerz@Darwin.Stanford.Edu)*

1 AVANTI Steering Committee. Analysis of HIV-1 clinical trials: statistical magic? *Lancet* 1999; **354**: 2061-64.