Perspective: Human Immunodeficiency Virus Type 1 (HIV-1) RNA End Points in HIV Clinical Trials: Issues in Interim Monitoring and Early Stopping


Department of Biostatistics, Harvard School of Public Health; Infectious Disease Division, Department of Medicine, Beth Israel–Deaconess Medical Center; Infectious Disease Division, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts; Department of Medicine, University of Miami School of Medicine, Miami, Florida

Due to the desire to both shorten the length and reduce the size of clinical trials in human immunodeficiency virus (HIV) disease, the use of surrogate end points such as HIV-1 RNA is becoming increasingly standard. While these end points may be reasonable surrogates for the clinical effectiveness of drugs, a key point in their use as trial end points is the definition of a relevant duration of antiviral response. This definition is often complicated by the desire to perform interim reviews of ongoing laboratory end point trials. Unlike clinical end point trials, in which early clinical response is generally indicative of longer-term follow-up, it is yet to be determined whether short-term viral response adequately predicts the long-term durability of that response.

There has been increasing movement in the human immunodeficiency virus (HIV) clinical trials community toward greater use of laboratory measures of biologic activity in the conduct and analysis of clinical trials of antiretroviral therapies [1, 2]. The use of laboratory markers in HIV clinical trials was historically motivated by a desire to shorten the duration and reduce the required number of patients of such clinical trials [3, 4] and more recently due to a better understanding of HIV dynamics [5–7]. The use of these laboratory markers implies that changes in specific measures of in vivo virologic or immunologic responses to antiretroviral therapies are associated with clinical benefit [8–10]. HIV-1 RNA copy number is such a laboratory marker [11, 12], and frequent HIV-1 RNA monitoring is rapidly becoming the standard of care in routine clinical practice; many clinical trials currently use HIV-1 RNA copy number as their primary end point. Interim monitoring or early evaluation of HIV clinical trials on the basis of laboratory markers is now considered ethically necessary by both patients and physicians and is thus essential if such clinical trials are to be conducted.

Change in HIV-1 RNA levels can be documented in a variety of ways when assessing the activity of antiretroviral therapies [13]. The usefulness of interim monitoring of HIV-1 RNA levels in antiretroviral clinical trials will depend critically on the manner in which the HIV-1 RNA end point is defined. Until recently, HIV-1 RNA end points have been defined in terms of the change from baseline level to a predetermined follow-up week, typically weeks 24 or 48. However, with the advent of more potent antiretroviral therapies, particularly protease inhibitor–containing regimens, recent interest has focused on end points defined in terms of the lower limit of quantification of the HIV-1 RNA assay, based on current theories about disease pathogenesis and viral resistance [14]. For clinical trials involving patients with initially detectable HIV-1 RNA levels, such end points are typically defined in terms of the proportion with undetectable levels at a given follow-up week (e.g., AIDS Clinical Trials Group [ACTG] 347). For clinical trials involving patients with initially undetectable HIV-1 RNA levels, the end point is typically defined as the proportion of patients who experience reemergence of detectable HIV-1 RNA by a given follow-up week (e.g., ACTG 343). Each of these choices of end points has unique problems when used for interim monitoring of HIV-1 RNA measurements.

We need to consider how interim monitoring should reflect the specific questions being asked by the study. The question of primary interest may be whether there exists a treatment difference with regard to initial drop in HIV-1 RNA level or with regard to the duration of suppression of HIV-1 RNA (durability). We also need to assess the potential reasons for early stopping of a clinical trial on the basis of interim monitoring. One reason to stop a study, or an arm of a study, is the failure of one or more treatments to affirm expectations (i.e., a prespecified minimum acceptable drop in HIV-1 RNA is not achieved). Another reason to stop a study is the obvious superiority of one treatment at a planned primary analysis time point or implied superiority based on results at an earlier time point. There are well-established statistical adjustments for multiple comparisons based on the subset of patients with available complete follow-up data to the prespecified week of interest. Additional theory needs to be developed for analyses that include shorter-term follow-up data.

For all classes of HIV-1 RNA end points, a key issue in interim monitoring of HIV-1 RNA levels is whether to use
only HIV-1 RNA levels at the primary follow-up time or to also make use of data at intermediate follow-up times. For example, if the primary end point is change in HIV-1 RNA level to week 48, an interim analysis could (1) be based solely on the HIV-1 RNA levels of participants who have completed week 48 follow-up at the time of the interim analysis or (2) use HIV-1 RNA levels at an intermediate follow-up week (e.g., week 12, week 24, or both) for those who have only completed follow-up to that point. The use of intermediate follow-up times has the obvious advantage of including more data in the interim analysis. However, the use of intermediate follow-up data in interim analyses requires a demonstration that intermediate HIV-1 RNA responses are predictive of later responses. For example, it may be reasonable to assume that a subject whose HIV-1 RNA is above the level of detection at week 24 continues to have detectable HIV-1 RNA at week 48. However, on the basis of current virologic assays and issues surrounding viral resistance, it may not always be reasonable to assume that a subject whose HIV-1 RNA is undetectable at week 24 will remain undetectable at week 48.

In clinical trials of less potent therapies or involving highly antiretroviral-experienced subjects, where change in HIV-1 RNA level rather than achievement of an undetectable level is deemed the appropriate end point, existing data suggest that intermediate HIV-1 RNA responses are indicative of later responses. For example, a metaanalysis of changes in HIV-1 RNA levels across three ACTG studies (175, 229, and 241) showed very high correlation between HIV-1 RNA responses at weeks 8 and 24 and those at week 48 (figure 1; [15]). This suggests that intermediate follow-up times may provide useful information when the primary end point is change in HIV-1 RNA level. Appropriate statistical methods must be developed for hypothesis testing based on interim monitoring under this situation.

It should be emphasized that a high correlation of weeks 8, 24, and 48 marker response data is a necessary but not sufficient condition for the use of week 8 and 24 marker data to imply week 48 marker outcomes. Figure 2 provides a hypothetical example in which RNA responses at weeks 24 and 48 are highly correlated, but the week 48 response may have less clinical relevance (HIV-1 RNA levels have returned to baseline) than the week 24 response (HIV-1 RNA levels are significantly reduced from baseline). Figure 3 provides a hypothetical example of a situation in which a treatment-induced marker effect at week 24 does not imply a treatment-induced marker effect at week 48. In this example, a study could have potentially been stopped at week 24 with claims of superiority of treatment A with regard to its effect on viral copy number; at week 48 it is seen that no superiority in fact exists. This highlights the need to use data from the primary end point, in conjunction with data from intermediate end points, when conducting such analyses. In addition, as the correlation of longitudinal marker responses and the nature of the biologic model is based on studies of less potent therapies, these findings must be evaluated cautiously with regard to their applicability to current clinical trials of more potent agents.

For potent therapies with which the appropriate end point is the proportion reaching and maintaining undetectable HIV-1 RNA levels, the extent to which HIV-1 RNA levels at intermediate follow-up times are indicative of those at later follow-up times is currently under investigation. In particular, the key point of interest in such clinical trials is the length of time for which HIV-1 RNA levels can be suppressed. Data from early follow-up times may not be informative about the duration of HIV-1 suppression; this relationship remains to be elucidated and may be drug-dependent. Thus, interim analyses based on data from intermediate follow-up time points in interim analyses have the potential to be misleading if the duration of suppression of HIV-1 RNA levels is the question of interest.

In contrast, for clinical trials using reemergence of detectable virus as an end point, reemergence by an intermediate follow-up time may imply continued levels of detectable virus at a later follow-up time (figure 4). Thus, intermediate follow-up times could prove useful in the interim monitoring of clinical trials utilizing time to virologic failure. It should be noted here that if the interval of time to undetectability (or reemergence) of HIV-1 RNA is the primary question of interest, traditional methods of sequential analysis are appropriate. However, in the case of an end point of viral undetectability, time-to-event analysis does not address the question of the durability of HIV-
1 RNA response, as patients are followed (for the analysis) only until the time that their level of HIV-1 RNA becomes undetectable. It is their time to viral reemergence that provides insight into treatment durability.

For any clinical trial that includes preplanned sequential monitoring of treatment effects, the choice of stopping rules is of special importance. This choice must be made in advance, with a balance of underlying assumptions and safeguards. Two commonly used stopping criteria are those of Pocock [16] and O’Brien and Fleming [17]. These choices ensure that repeated tests of treatment effects will be valid at a prespecified level of significance. In the case of the Pocock stopping rule, the $P$ value required to reject the null hypothesis remains constant over all interim analysis times, whereas the O’Brien-Fleming rule requires much more substantiated treatment effects at early interim analyses, but a significance level at the final analysis that is close to the overall $P$ value (generally .05). The Pocock rule allows stopping at an earlier analysis time with less extreme treatment differences but requires larger treatment differences at the end of the study than does the O’Brien-Fleming rule to achieve the desired level of significance. The O’Brien-Fleming rule tends to continue studies to their planned termination.

When early follow-up HIV-1 RNA levels are used to impute levels at a later follow-up time, additional problems must be considered. There is a need not only to preserve the desired overall significance level through methods of sequential monitoring, but also to be confident that there is sufficient evidence to ensure that the imputation is valid. In this setting, early stopping has an additional drawback; not only is less information gathered about the durability of treatment effects, but also fewer data will be available to validate the imputation. Statistical theory that accounts for the imputed nature of the data is currently being developed for hypothesis testing of the primary end point.

Table 1 provides examples under which imputation may be more or less valid. In the first example, patients show durability in their HIV-1 RNA detectability; the detectability observed at week 24 always properly implies the detectability at week 48. Under these circumstances, while the criteria for significance

Figure 2. Correlation of weeks 24 and 48; week 48 is clinically irrelevant. Data are changes in HIV-1 RNA level with 2 treatments (A and B).

Figure 3. Early treatment difference is not equal to late treatment difference. Data are changes in HIV-1 RNA level with 2 treatments (A and B).
must be adjusted for multiple comparisons (multiple interim analyses), no further penalty for premature decision need be imposed. In the second example, however, week 24 does not always predict week 48. In this situation, an extra penalty should be imposed due to the uncertain relationship between the early and later outcomes. Only if there is sufficient reason to assume that the first scenario is true is the Pocock rule reasonable. Under the second scenario, the more stringent criteria for early stopping provided by the O’Brien-Fleming rule is likely to be preferred. Continuation of the clinical trial may also provide the information necessary to further understand the relationship between early and late HIV-1 RNA levels. This decision must be weighed carefully, taking into account a variety of complicating issues.

One such issue is due to the likelihood that response to new treatments is dependent not only on the duration of previous treatment, but also the specific nature of previous treatments. It is reasonable to assume that certain treatment regimens may induce viral resistance (either complete or partial) to subsequent regimens [18]. For this reason, it is our view that interim values of HIV-1 RNA do not capture all of the information needed to plan long-term treatment strategies. We also need to gather data on the impact of specific treatment regimens on the success of specific subsequent regimens. Only by considering this information and data on the toxicity profiles of different regimens can a well-informed long-term treatment strategy be developed.

In summary, we believe that developing appropriate statistical methods for early termination of HIV clinical trials based on preliminary HIV-1 RNA findings is of critical importance. It is essential that interim analyses reflect the primary end point and that early study termination must be based on the evaluation, either directly or by implication, of the primary end point. This is particularly important for HIV clinical trials in which durability of therapeutic response is the primary study question, since great care must be taken to distinguish acute treatment-induced effects from their durability.

We believe that in some carefully defined circumstances, interim HIV-1 RNA data in conjunction with available complete follow-up data could form the basis for early termination of a clinical trial. If there are already conclusive data from other clinical trials about the relationship between HIV-1 RNA response at earlier time points and at the primary end point, then basing interim analyses on earlier time points may be justified as long as the additional uncertainty inherent in projecting later responses from early data is taken properly into account. If the information about this relationship is unavailable from an external source, then accumulating data from patients who have been followed long enough that the primary end point has been observed must be used to assess the ability of HIV-1 RNA response at earlier time points to predict HIV-1 RNA response at the time of the primary end point.

While it is crucial that patients not continue to receive inferior treatments, we would like to emphasize that this does not preclude gathering important information about long-term follow-up, including but not limited to HIV-1 RNA and clinical outcomes, for patients who are maintaining their assigned treatment and for those who have switched to other treatments. Only through long and committed follow-up can the intricate

**Table 1.** Examples of imputation of long-term HIV-1 RNA results from short-term data.

<table>
<thead>
<tr>
<th>Imputation</th>
<th>Detection at</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Week 12</td>
<td>Week 24</td>
<td>Week 48</td>
</tr>
<tr>
<td>Reliable</td>
<td>Above/below</td>
<td>Below</td>
<td>Below</td>
</tr>
<tr>
<td>Below</td>
<td>Above</td>
<td>Above</td>
<td></td>
</tr>
<tr>
<td>Above</td>
<td>Above</td>
<td>Above</td>
<td></td>
</tr>
<tr>
<td>Unreliable</td>
<td>Above/below</td>
<td>Below</td>
<td>Below/above</td>
</tr>
<tr>
<td>Below</td>
<td>Above</td>
<td>Above</td>
<td></td>
</tr>
<tr>
<td>Above</td>
<td>Above</td>
<td>Above/below</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.** Trial investigating reemergence of detectable virus. Data are changes in HIV-1 RNA level with 2 treatments (A and B).
relationship between treatment-mediated marker responses and clinical outcome be elucidated.

References