Published in final edited form as:

Control Clin Trials. 2001 April ; 22(2): 142–159.

ACTG (AIDS Clinical Trials Group) 384: A Strategy Trial Comparing Consecutive Treatments for HIV-1

Laura M. Smeaton, MS, Victor DeGruttola, ScD, Gregory K. Robbins, MD, MPH, and Robert W. Shafer, MD

Center for Biostatistics in AIDS Research, Harvard School of Public Health, Boston, Massachusetts (L.M.S., V.D.), Massachusetts General Hospital, Boston, Massachusetts (G.K.R.), and Stanford University Medical Center, Stanford, California (R.W.S.)

Abstract

ACTG (AIDS Clinical Trials Group) 384 is designed to evaluate different strategies for antiretroviral treatment in HIV-1-infected individuals with no previous exposure to antiretroviral treatment. The study is a randomized, partially double-blinded, controlled trial with 980 subjects at 81 centers in the United States and Italy. The study has a factorial design that addresses the following scientific questions: (1) Does the best initial choice of therapy include both a protease inhibitor (PI) and non-nucleoside reverse transcriptase inhibitor (NNRTI) in a four-drug combination with nucleoside analogue (NRTI) drugs, or should these agents be used sequentially in three-drug combinations?; (2) Which sequence is best in a three-drug regimen—PI followed by NNRTI or NNRTI followed by PI?; (3) Which is the best sequence of dual NRTI combinations—zidovudine plus lamivudine followed by didanosine plus stavudine, or the converse? Subjects in the three-drug combination arms are offered a salvage regimen after failure of their second regimen; subjects in the four-drug combination arm are offered a salvage regimen after failure of their first regimen. The primary endpoint of the study is the time until salvage; secondary endpoints include time to virological failure and time to toxicity-related discontinuation of therapy. A Division of AIDS Data and Safety Monitoring Board will review the trial for safety and efficacy.

Keywords

Clinical trials design; phase III study; randomized; controlled; multicenter; double-blind trial; AIDS

INTRODUCTION

With 14 HIV antiretrovirals in three drug classes (six nucleoside reverse transcriptase inhibitors [NRTIs], five protease inhibitors [PIs], and three non-nucleoside reverse transcriptase inhibitors [NNRTIs]) approved by the Food and Drug Administration (FDA), there are an enormous number of possible choices for initial treatment of HIV-1 in infected

individuals. The goals of any initial treatment strategy are long-term suppression of plasma HIV-1 RNA with limited adverse effects and toxicity, as well as preservation of future treatment options for subjects who fail initial treatment. Numerous clinical trials have demonstrated the superiority of three or more antiretrovirals in combination over single agents and dual-agent combinations [1–7]. Current recommendations from the International AIDS Society are that three or four drugs from two or three of the drug classes are generally required to suppress plasma HIV-1 RNA levels [8]. ACTG (AIDS Clinical Trials Group) 384 addresses the question of identifying the best sequence of three- and four-drug combinations.

Two competing approaches for initial treatment are to use four agents, chosen from all three drug classes, or to use only three drugs from two classes, thus preserving another class for subsequent therapy. The former approach may have more potency but exposes the individual to additional toxicity and more problems for adherence compared to the three-drug approach. When a subject fails a four-drug combination, he or she may not respond as well to subsequent (salvage) regimens because of the possible development of cross-resistance to drugs in all three classes [9–13]. A possible shortcoming of drug class-sparing strategies is a greater chance of incomplete suppression, or more rapid rebound, of plasma HIV-1 RNA. Low levels of HIV-1 RNA replication, even if unmeasurable, may permit selection of mutations of HIV-1 that confer drug resistance and thereby result in early failure of the drug combination [4, 6, 14–18]. Therefore, the choice of first antiretroviral therapy has a strong impact on the entire future of the individual's treatment options. Comparing the outcomes of various sequences of therapy is necessary to understand the long-term effects of choice of initial treatment strategy.

To address these questions, the ACTG developed a randomized, controlled clinical trial to compare different strategies for initial antiretroviral treatment in HIV-1-infected individuals. In our context, a treatment strategy is defined as an approach that includes one or more combination regimens in succession. Therefore, it involves not only the initial selection of treatment, but also selection of follow-up regimens for those subjects who fail the initial treatment. ACTG 384 is designed to study sequences of two or three combination regimens to determine which are tolerable and provide durable suppression of HIV-1 over a 2- to 3-year follow-up period. Nine hundred eighty subjects at 81 sites across the United States and Italy were accrued to the trial in approximately 1 year; the study is designed to continue follow-up for another 2 years after completion of accrual. The primary endpoint of the study is the time to failure of the treatment strategy (discussed later in this article).

STUDY DESIGN

ACTG 384 has six treatment arms that comprise a 2×3 factorial array; each treatment arm consists of a sequence of combination regimens. All of the individual drugs in the first two sequences of regimens have been approved by the FDA for treatment of HIV. The first factor of the design, the nucleoside (or NRTI) factor, has two levels defined by the initial choice: didanosine plus stavudine (ddI plus d4T) or zidovudine plus lamivudine (which is given in a single ZDV/3TC tablet). The second factor, choice of initial antiretroviral drug classes, has three levels: initial NNRTI (efavirenz [EFV]), initial PI (nelfinavir [NFV]), and

initial PI plus NNRTI (NFV plus EFV). The design may also be viewed as a $2 \times 2 \times 2$ factorial with two missing cells [19]. (See Tables 1 and 2 for overview of study design.)

Arms A, B, C, and D each include a sequence of two three-drug combinations followed by a salvage regimen; each of these four arms is a drug class-sparing strategy. Arms A and C initially include EFV and therefore may be viewed as a PI-sparing strategy. Subjects in these two arms are offered NFV as part of their second regimen, after failure of the first regimen. Arms B and D have the reverse strategy, initially including NFV, and delaying use of the NNRTI drug class until failure of the first regimen. Arms E and F start treatment with four-drug combinations incorporating both NFV and EFV. Because no class of drugs has been spared, the second combination in both of these arms is the salvage regimen. Arms A, B, and E all start with ddI plus d4T, and arms A and B follow with ZDV plus 3TC (the salvage for arm E may include this combination as well). Arms C, D, and F employ the converse sequence of NRTIs. When a subject progresses to a new regimen combination in the sequence, at least three new drugs are introduced. The treatment arms are partially double-blinded; the NRTIs are provided open label, whereas the EFV and NFV are double-blinded via matching placebos.

All drugs in the salvage regimens (after the primary study endpoint has been reached) are provided open label. ACTG 384 defines a salvage regimen as a treatment combination initiated after exposure to all three major antiretroviral drug classes. As the optimal salvage regimen was not known at the outset of the trial, the specific drugs that comprise the salvage regimens may change over the course of the trial. Initially, the salvage regimens consisted of a highly potent dual-PI combination (amprenavir plus indinavir), and either two previously unused NRTIs for subjects in arms E or F or the combination of ddI with a compound that increases its NRTI activity intracellularly (hydroxyurea) for arms A through D. In July 2000, the study was amended to include updated salvage regimens. Subjects in arms A through D, along with their physicians, may choose from six different salvage regimens (a PI combination of either amprenavir plus ritonavir, or indinavir plus ritonavir, plus one of the following: ddI plus hydroxyurea, abacavir plus ZDV/3TC, or abacavir plus d4T plus 3TC). Subjects in arms E or F choose between the two PI combinations above, matched with the NRTI combination not used in their previous regimen.

The primary objectives of the study can be formally stated as follows: First, is a drug combination that starts with both a PI and an NNRTI preferable to a sequence starting with either a PI-sparing combination or an NNRTI-sparing combination? (See Table 3.) This objective compares a four-drug combination incorporating drugs from all three antiretroviral drug classes (NRTI, PI, and NNRTI) with two three-drug combinations that each spare a drug class (either PI or NNRTI) for later use. Second, with three-drug combinations, is it better to start with a PI (followed by an NNRTI) or an NNRTI (followed by a PI)? (See Table 4.) This objective compares the optimal sequence of two drug classes, PI and NNRTI, when used as part of a drug class-sparing strategy. Third, what is the best sequence of NRTIs in a three- or four-drug combination: starting with ZDV plus 3TC or starting with ddI plus d4T? (See Table 5.)

Because of the large number of questions regarding choice of initial therapy and the limited number of subjects available to participate in such research, it is important to use efficient study designs. Factorial study designs have a number of advantages in this setting [20]. First, they allow the simultaneous evaluation of multiple objectives within a single trial. Each of the three main scientific objectives of the trial may be translated into one or more statistical comparisons among the levels of a factor. For example, the third main objective of the trial is to compare the sequence of NRTI dual combinations, which is achieved by comparing arms A and B versus arms C and D. (A secondary comparison is arms A, B, and E versus arms C, D, and F.) Second, a factorial design allows assessment of interaction between the factors, i.e., the degree to which the level of one factor influences the effect of the other. A hypothetical example of an interaction in this context is the following: one sequence of dual NRTIs is better for a PI-sparing strategy, but worse for an NNRTI-sparing strategy. Third, if additivity holds (in other words, if there is no significant interaction between the factors), then a factorial design is more efficient than multiple one-factor trials, providing more power for comparisons of the main effects. The reason for this gain in efficiency is that evaluation of one factor involves pooling over the other factor; such pooling increases the effective sample size of each comparison [21, 22]. For instance, to compare the strategies of initial use of a PI versus initial use of an NNRTI (the second study objective), pooling over the NRTI factor means that arms B and D are combined and compared with arms A and C combined.

STUDY POPULATION

The target population for ACTG 384 is antiretroviral-naïve individuals with documented HIV-1 infection. They may be newly infected, newly diagnosed, or other previously untreated individuals. The eligibility requirements allow no more than 6 days of exposure to antiretrovirals for HIV-1. This requirement aims to minimize chances of drug resistance at baseline and ensure that the sequences of drug combinations provided in the study are potentially viable for all subjects. Eligibility criteria include a plasma HIV-1 RNA of at least 500 copies/mL, so that current assays can measure the initial potency of these treatment combinations. Other major eligibility requirements include the following: age of at least 13 years, acceptable laboratory values, willingness to practice birth control (because of possible teratogenicity with study medications), not pregnant or breast feeding, absence of contraindicated concomitant medications, and informed consent.

The baseline characteristics of the subjects who enrolled in ACTG 384 are given in Table 6. Of the 980 subjects enrolled, 181 (18%) are female. The median age is 36, and 11% of subjects are at least 50 years of age. The study population is ethnically diverse; 53% of subjects are not white. Nine percent of subjects report current or previous intravenous drug use. The median CD4 cell count is 277.5 cells/mm³. Twenty-five percent of subjects have a baseline CD4 cell count fewer than 100 cells/mm³, and 18% of subjects have a CD4 cell count greater than 500 cells/mm³. The median plasma HIV-1 RNA is 86,868 (or 4.94 log₁₀ transformed) copies/mL. One percent of subjects had plasma HIV-1 RNA fewer than 500 copies/mL after screening above 500 copies/mL. Eighteen percent of subjects have at least 500,000 copies/mL of HIV-1 RNA in plasma.

CHOICE OF ENDPOINT

The primary study endpoint is time to salvage regimen, i.e., the time to completion of the first two treatment regimens for arms A through D, and time to completion of the first regimen in arms E and F (as the second treatment regimen is salvage). For those subjects who refuse either further study treatment or study evaluations, or who are lost or die before completing the regimen sequence, the endpoint is defined as time to permanently stopping study treatment. Subjects fail their current treatment and progress to the next for reasons that include adherence, toxicity, intolerance, and poor virological response as measured by plasma HIV-1 RNA. Therefore, the endpoint combines information about toxicity and virological response.

While toxicity and virological efficacy are analyzed separately in the investigation of secondary endpoints, the primary endpoint combines them. One reason for this choice is that plasma HIV-1 RNA measures both virological efficacy as well as toxicity and intolerance because the latter two affect regimen adherence. As the primary goal of successful antiretroviral therapy is for subjects to remain on a tolerable regimen that suppresses the virus, discontinuation of the regimen for any reason constitutes a failure of the regimen. The cost of such a failure may be high, because exposure to these drugs may reduce future options. Secondary endpoints of the study are defined to permit separate investigation of efficacy and safety. While analysis of these endpoints can help disentangle the reasons for treatment failure to some degree, toxicity and efficacy remain closely tied. In fact, characterizing the cause of a regimen failure can be somewhat arbitrary. For example, subjects with low-level toxicity may not adhere to their regimens and therefore may experience virological rebound. It is often difficult to assess whether the failure results strictly from either toxicity or lack of efficacy when both precede a regimen failure. In recognition that it is not always possible to distinguish the original, or causative, event in a regimen failure, the study team chose to focus on an endpoint that most closely reflects the goal of antiretroviral treatment. That goal is for a subject to be able to remain on a regimen that is tolerable and that provides durable suppression. The study protocol specifies that all subjects are to be followed for the duration of the study, regardless of treatment (or endpoint) status. Continued follow-up after treatment discontinuation or meeting the primary study endpoint is important so that subjects may be evaluated for secondary endpoints, such as time to virological failure. It is possible that a study design that incorporates changes in treatment may have an increased risk of dropout following a treatment change whether or not the treatment change meets a primary study endpoint. Because follow-up is specified beyond the primary study endpoint, the composite primary endpoint should not seriously compromise the ability to analyze each of the components (specifically, virological failure and toxicity) of regimen failure separately.

CRITERIA FOR TREATMENT FAILURE

To minimize differences in subject management that could affect the primary endpoint, the study protocol defines guidelines for determining when a drug regimen has failed. For example, the study document defines levels of toxicity and intolerance that require dose reductions, temporary holds, rechallenging (when appropriate), and finally abandoning that

regimen for the next. When medically appropriate, adjustments such as dose reduction or limited, single-agent substitution (only 3TC for ddI or d4T for ZDV are allowed) are considered before declaring the regimen a failure.

The protocol specifies four criteria for virological failure that were picked to mirror clinical practice at the time the study was designed. The first is a lack of initial virological potency of the treatment regimen, defined by a plasma HIV-1 RNA level at week 8 that does not drop at least 1 log₁₀ copies/mL from baseline level and is more than 200 copies/mL. Second is early rebound in the first 24 weeks, defined by a rise in plasma HIV-1 RNA more than 1 log₁₀ from the subject's nadir level on the current regimen and that is more than 2000 copies/mL. Third is a loss of virological suppression characterized by a measure above 200 copies/mL after two consecutive values below 200 copies/mL. Fourth is a failure to suppress or a loss of virological suppression, which occurs when plasma HIV-1 RNA is above 200 copies/mL on or after 24 weeks on a regimen. The assay used to measure HIV-1 RNA in plasma is the Roche Ultrasensitive Assay, which has a lower limit of quantification of 50 copies/mL. Assays are performed at a central laboratory (to reduce interlab variability) in real time, and results are reported back to the subject's site within 1 to 2 weeks of specimen draw. All virological regimen failures require a confirmation HIV-1 RNA result within a 6-week period. The reason for requiring confirmation is that a single rise in HIV-1 RNA due to short drug holds (or subject-initiated drug holiday), intercurrent illness, or measurement error may falsely indicate that the regimen is no longer effective. In addition, to match clinical practice, a second confirmation measure is allowed before declaring a regimen a failure whenever the first confirmation value is at least 1 log₁₀ lower than the first HIV-1 RNA measure meeting one of the failure criteria.

Subjects are required, as part of the study protocol, to progress from one treatment regimen to another as they meet the study-defined criteria for switching regimens. This information was included in the template informed consent document so that potential subjects would be informed of these requirements of the study before agreeing to participate. If a subject meets the criteria defining a regimen failure and refuses to switch, the study protocol states that the subject should be permanently discontinued from all study medications and continue to be followed on study after treatment discontinuation. However, there have been a handful of situations in which this rule has not been strictly followed. In one type of situation, a subject has only just met the criteria for regimen switch. The team recognizes that the threshold levels defined in the criteria (such as 200 copies/mL) have some level of arbitrariness. Therefore, the team has allowed a handful of subjects in "borderline cases" to continue the current regimen after meeting one of the criteria to switch to see if the subject's HIV-1 RNA value drops below 200 copies/mL at the next measurement. Another type of situation has been when a subject has met one of the virological criteria for regimen switch only after a mandated temporary hold of study treatment for a condition unrelated to treatment (such as pregnancy or temporary incarceration). The core study team has monitored all these situations very closely to make sure that these cases are exceptions, rather than the rule. Frequent communication between the team and sites, along with criteria that admit a small degree of flexibility (such as allowing for a second confirmation when plasma HIV-1 is falling or allowing up to 6 weeks to draw a confirmation sample) have helped improve adherence to the regimen-switching criteria without compromising the study.

RANDOMIZATION AND SAMPLE SIZE/POWER

ACTG 384 was originally designed to accrue a total of 800 subjects over a period of about 1 year to the six arms of the study. The allocation ratio was 5:5:5:5:6:6 resulting in sample sizes of 125 subjects each in arms A, B, C, and D and 150 subjects each in arms E and F. The randomization was implemented via permuted blocks in two stages. In the first stage, the total block size for the second stage was determined by a 4:1 allocation to block size of 6 or 8, respectively. Then, in the second stage, those subjects randomized to a block size of 6 had equal allocation (1:1:1:1:1:1) to each of the six arms, and those randomized to a block size of 8 had a 1:1:1:1:2:2 allocation to arms A through F, respectively. Alternatively, one could randomize using permuted blocks of size 32. However, the two-stage randomization controls the balance among study arms at any point during the realization of the randomization, so that halting accrual in the middle of a block may lead to less chance imbalance among the arms compared to a block size of 32. Concurrent randomization to the treatment arms was stratified by screening plasma HIV-1 RNA (<10,000 copies/mL, 10,000–100,000 copies/mL, and >100,000 copies/mL) and country of participation (United States or Italy). Stratification by HIV-1 RNA was used because baseline HIV-1 RNA has been shown to be a prognostic factor for virological response to potent combination regimens. The allocation among the arms of the study was chosen to maximize the power with the least number of subjects required for each of the primary study objectives.

We estimated the necessary sample size for this study by modeling the probability (p_1) of viral load falling to below quantifiable levels after the initial regimen, the proportion (r_1) of subjects who will still be below detectable levels at 1 year, the probability (p_2) that viral load will fall below detectable levels on the second regimen (after failure on the first regimen), and the proportion (r_2) of subjects given the second regimen who have undetectable viral load 1 year later. For the arm beginning with ddI plus d4T plus EFV, we assumed $p_1 = 0.68$, $r_1 = 0.50$, $p_2 = 0.70$, and $r_2 = 0.50$. For the arm beginning with ddI plus d4T plus NFV, we assumed $p_1 = 0.90$, $r_1 = 0.75$, $p_2 = 0.60$, and $r_2 = 0.40$. These assumptions lead to predictions that among subjects in the arm that begins with ddI plus d4T plus EFV, 64% (52%) will complete 2 (3) years before failing on their second regimen. For the arm beginning with ddI plus d4T plus NFV, the estimated probabilities were 76% and 66%. We used the 3-year survival rates to obtain average hazard rates of 0.218 and 0.139, respectively, and a hazard ratio of 1.57. For analyses of primary endpoint, we assumed that arms E and F would have a hazard rate similar to those of the better three-drug regimen above, 0.139.

Power is then calculated assuming an exponential failure time model, adjusted for uniform accrual of 1 year and follow-up of 2 years [23]. With these assumptions and an increased sample size due to overenrollment, the estimated power for each of the primary study objectives is the following: the first objective of four-drug strategy versus three-drug strategies (with adjusted type I error of 0.025), 89% power (310 versus 360 subjects), and 92% power for the other two main objectives (310 versus 310 subjects).

The primary endpoint occurs when subjects permanently stop study treatment, even if it is before completing the sequence of regimens requiring initiation of salvage. These early

discontinuations are due to a number of different causes, such as loss to follow-up or subject withdrawal. Some dropout may be related to the effects of treatment, such as low-grade toxicity/intolerance or clinical progression, and therefore it is appropriate to include these as part of the primary endpoint. However, some losses to follow-up are clearly unrelated to the study treatment (e.g., when a subject moves away or becomes incarcerated) and can be considered noninformative. It would be ideal to be able to distinguish reliably between informative and noninformative treatment discontinuations and include only the former in the primary study endpoint. In practice, though, this identification is not always possible. Therefore, the three main options for classifying treatment discontinuations are the following: count as endpoint (and consider all as informative), censor (and consider all as noninformative), or model [24]. We chose the first of these as being best for the primary analysis of this study. To examine the extent to which noninformative dropout could attenuate the primary treatment comparisons, and therefore reduce power, we performed the following calculations.

Under the alternative hypothesis for the second and third primary objectives, the expected number of endpoints (for a hazard ratio of 1.57 with 92% power) is 221 (or 91 in the better group and 130 in the worse group). The average rate of loss to follow-up over all ACTG studies is 5.17 per 100 subject-years. Using this rate, we project that approximately 126 subjects in ACTG 384 will be lost to follow-up during the course of the trial. Considering all of these losses to be noninformative (a conservative assumption, in this case), we distribute these endpoints among the groups proportionally by treatment-arm size. Assuming the alternative hypothesis applies to the subset of subjects not lost to follow-up, we would expect 80 endpoints in the better group and 114 in the worse group. With these assumptions, we observe 120 endpoints plus 154 endpoints, or a total of 274 endpoints, because of the addition of 40 losses to follow-up in each group. The resulting power associated with the observed number of endpoints is 80%, which represents a 12% loss of power from the original design of the study. This same reasoning can be applied to the other (first) primary study comparison to estimate a drop of power from 89% to 77%. We believe that these are conservative estimates for loss of power because the calculations treat all losses to follow-up as primary study endpoints. In reality, we expect that a portion of subjects who become lost will have previously met the primary study endpoint based on the regimen-switching criteria. Therefore, it does not appear that the power of the study was seriously compromised by this design feature.

STATISTICAL ANALYSES OF THE PRIMARY STUDY ENDPOINT

Each of the three primary objectives will be tested by a pairwise comparison between the two levels of the factor of interest, using the log-rank test. Bonferroni adjustment of type I error for multiple comparisons will be made for the first objective (four- versus three-drug strategies), as two pairwise comparisons (among the three levels of the factor) are made in order to answer this one objective [25]. The Mantel-Haenszel form (weights = 1) of the log-rank test will be used, as it is most powerful (compared to others in the Tarone-Ware class of tests), under the assumption of proportional hazards [26, 27]. The log-rank test will be stratified by the levels of the factor being pooled, as well as baseline stratification factors. Assessment of any interaction between the two factors (and between the main effects and

the stratification factors), will be made using Cox proportional hazard modeling [28]. Secondary analyses include evaluating toxicity and virological failure separately, using the same survival analysis methods as for the primary endpoint.

CONDUCT AND MONITORING

ACTG 384 is being conducted at 32 main ACTG sites (plus 26 subunits) across the United States and at 23 sites in Italy. The study will be reviewed by the Data and Safety Monitoring Board of the Division of AIDS on an ongoing basis. The timing of the interim reviews is partially determined by the number of primary study endpoints, with at least one review per year and three reviews planned in total. Peto adjustment of the significance level for sequential monitoring will be employed [29]. The study opened to accrual in October 1998 and closed to accrual at the end of November 1999. Therefore, the trial is currently planned to close to follow-up (barring any early stopping based on planned interim reviews) in late 2001.

Data collection in ACTG 384 is extensive. After a screening visit to determine eligibility and obtain written informed consent, subjects were evaluated at entry to the study (before dispensation of study treatment) to determine baseline characteristics. Table 6 shows a number of baseline parameters collected. In addition, a clinical assessment with diagnoses and signs and symptoms was made, and adherence prediction and quality of life questionnaires were administered. During follow-up, subjects are evaluated every 4 weeks during the first 6 months on each study regimen, and then seen every 8 weeks thereafter. At every study visit, a targeted physical exam for diagnoses and a signs and symptoms assessment take place. Also, blood is drawn for real-time plasma HIV-1 RNA testing and for specimen storage for future virological (such as resistance) and immunological tests. Every 8 weeks the following laboratory tests are performed: hematology, blood chemistry, liver function tests, and pregnancy tests (for appropriate subjects). CD4/CD8 cell counts and flow cytometric analyses are performed every 8 weeks during the first year on a study regimen and every 16 weeks thereafter. At specified visits, adherence and quality-of-life questionnaires are administered. Finally, twice during each pre-salvage regimen, population pharmacokinetics samples are collected. Additional targeted evaluations are necessary for subjects enrolled in any of the five substudies of ACTG 384. Initiation of a new treatment regimen resets the evaluation schedule, so that subjects are more closely monitored during the first 6 months of taking new drugs.

Data are collected on case-report forms, keyed remotely at the site, and then electronically transmitted to the central database daily. Every evening, submitted data are checked by a logical checks computer program for internal (within-form) completeness, accuracy, and consistency. For example, accuracy of the subject identifier code is checked by cross-referencing with a master list of all subjects on the study. Data managers match coded values against text descriptions and resolve discrepancies via an electronically tracked querying system using e-mail. Currentness and completeness of data are also assessed with a monthly delinquency program that anticipates what data are expected based on each subject's key milestone dates (such as randomization, off-study, or death date) and the data collection schedule. Cross-form and longitudinal data checking are performed by both the

data manager and statisticians. Key data relating to the primary study endpoint, such as plasma HIV-1 RNA results, are monitored frequently and intensely. As these data are received directly from the testing laboratory in real time and decisions to change regimens are based on these results, plasma HIV-1 RNA data are managed on a weekly basis by a team consisting of the data manager, the statistician, and the laboratory data coordinator. Plasma HIV-1 RNA results undergo over 20 logical checks including verification of the validity of the specimen date, subject identification number, and RNA result. Discrepancies and possible errors are queried in real time.

SUBSTUDIES

ACTG 384 has five substudies to address focused questions on smaller cohorts of subjects. Each substudy enrolls consenting and qualifying subjects from the main study population.

The largest of these is a metabolic substudy that is designed to examine glucose metabolism, lipid disorders, and body composition changes. The primary objectives of the substudy are to compare insulin resistance, cholesterol, and triglyceride metabolism between subjects initially receiving a PI compared to subjects initially receiving an NNRTI. The accrual goal of 345 subjects was almost met, with 337 subjects enrolling in the substudy. Special evaluations for this substudy include the following: blood serum/plasma and urine collection, standardized body weight, body composition (bioelectrical impedance analysis, when possible), body measurements, body image questionnaire, optional food intake diaries, optional regional body composition by dual-energy X-ray absorptiometry, and optional oral glucose tolerance test.

A pharmacology substudy will examine the pharmacokinetics of NFV and EFV by means of 12-hour intensive pharmacokinetic studies. These intensive substudies collect nine blood samples (and two peripheral blood mononuclear cell specimens) over a 12-hour period and are performed 4 and 32 weeks after starting the relevant study regimen. This substudy will also consider how sequential therapy with the two nucleoside analogue combinations influences the intracellular phosphorylation patterns of ZDV, d4T, and 3TC. This substudy has a total accrual goal of 100 subjects and has currently enrolled 67 subjects. Subjects may join when beginning the second nonsalvage regimen (from the three-drug arms only); therefore, this substudy is still accruing.

An immunology substudy enrolled 77 (out of a goal of 90) subjects for the purposes of describing changes in naïve (CD45RA/CD62L) and memory (CD45RO) phenotypic markers of CD4/CD8 lymphocytes and their relationship to apoptosis and HIV viremia. This substudy's other primary objectives include describing changes in markers of immune activation (CD38 and DR) of CD4/CD8 lymphocytes, describing changes in plasma cytokines and soluble activation markers (TNF-alpha, sTNFR II, sIL-2R), and assessing lymphocyte proliferative responses to nonspecific and specific recall antigens. Subjects in this substudy have a baseline computed tomography chest scan. Every 16–24 weeks, advanced flow cytometry for T-cell activation and apoptosis and lymphocyte proliferative assays are performed in real time, and samples (cells, serum, and plasma) are collected.

There are two similar adherence substudies nested within ACTG 384. The smaller is a 24-week adherence substudy conducted only at five main trial units. Participants in the substudy are randomized between standard of care and additional telephone calls from a central location (not local study personnel). The objective of this substudy is to examine how additional telephone support affects subject adherence to combination antiretroviral therapy. A total of 109 subjects (out of a goal of 150) enrolled in this substudy. Subjects have an electronic monitoring system (MEMS cap) fitted to the NFV/placebo bottle. These bottle caps are evaluated and pills are counted at each visit during the first 24 weeks on study.

The second adherence substudy also randomizes subjects equally between standard of care and additional structured telephone support (this time by local study personnel, and less frequently than the previous substudy). The primary objective of this substudy is to evaluate the effect of this intervention on adherence, quality of life, dosage reductions, initial regimen failures, and virological outcomes. This substudy continues for the length of the main study and enrolled 335 (out of a goal of 350) subjects. All the evaluations relevant to this substudy (adherence, quality-of-life questionnaires) are collected under the auspices of the main study.

DISCUSSION

ACTG 384 is intended to provide guidance about the use of sequential therapies in treating antiretroviral-naïve patients with HIV infection. The challenges in this study result from the need to specify when and how patients should switch therapies, but still permit enough flexibility so that subjects will be willing to comply. As a result, the protocol must make sure that all of the specified treatments remain appealing to subjects and consistent with ongoing standards of care throughout the course of the study. Following subjects beyond their first regimen to their second and sometimes third regimen in a controlled fashion allows the study to compare the longer-term outcomes of these strategies as implemented in sequences of combination treatment regimens.

Perhaps the most important design feature of any study is the choice of primary endpoint. To design a study based on clinical endpoints would be futile in the current research environment, because occurrence of clinical events in this population is low. As a result, the ACTG has developed a study using a marker endpoint, with an added feature that all subjects will be followed for long-term outcomes after the conclusion of the study, via the ACTG Longitudinally Linked Randomized Trial. In practice, this is the only way to answer long-term questions about the clinical consequences of different ways to initiate therapy. The question of whether the marker should be a measure of viral burden, of immune function, or a composite of both is a primary concern in all areas of HIV/AIDS research. But this question can only be answered with data from studies of long-term clinical outcome, which is exactly the kind of data that ACTG 384 will ultimately produce. These data are required to evaluate the “surrogacy” of different marker-based endpoints, i.e., the degree to which treatment effects on the marker predict treatment effects on a clinical outcome.

Until there is research on the surrogacy of different endpoints, the selection of endpoints must be based on current clinical beliefs about the value and meaning of HIV disease markers. An important belief that strongly affects the design of HIV trials is that allowing patients with detectable virus to remain on the same treatment regimen will lead to the development of increasingly resistant virus. Moreover, it is postulated that the emergent virus will be resistant not only to the treatments to which the individual has been exposed but also to other members of the drug class. Such beliefs (and there is some evidence for them [30]) have led many clinicians to change therapy as soon as viral rebound is observed, on the theory that potent salvage is still possible for individuals who have not yet had extensive antiretroviral exposure. For this reason, the time of “regimen completion” is a natural and meaningful endpoint. Combining toxicity and efficacy information into the primary study endpoint is controversial because the composite endpoint does not distinguish toxicity/safety from efficacy information. It does have the advantage, however, of a close relationship to actual clinical management. Furthermore, separate analyses of time to “purely” virological failure and of toxicity will also be undertaken.

There have been a number of challenges in implementing ACTG 384. These have typically focused on educating sites (physicians and research nurses) about how this trial design is fundamentally different from many of the other ACTG studies they conduct, because the endpoint is tied to clinical management (via regimen completion). In many trials, after subjects experience toxicity or fail the study regimen, they discontinue study treatment (and may start medications prescribed outside the study), but they are still followed for the primary study endpoint. Because ACTG 384 specifies subsequent regimens until after the occurrence of the primary study endpoint, this choice is taken away from the health professional and subject. The team has worked very hard to educate and justify the rationale for removing this flexibility. At first, one of the challenges was implementing complex criteria for regimen failure. Efforts to help implement the criteria have included site education, such as a specialized RNA worksheet to aid sites in determining if any of the virological criteria have been met. Another mechanism has been very frequent monitoring by the team, particularly through weekly RNA reports that remind sites when criteria have been met or when confirmation samples need to be drawn. And finally, implementation has required close communication between the sites and the study chairs.

One of the most frequent challenges with defining regimen failure has been adherence. Sometimes, clinicians will contest a regimen failure, claiming that the elevated viral loads meeting the criteria are due to the subject not taking the medications. Therefore, the clinician postulates that this is not a “true” regimen failure requiring a change in treatment regimen. The team has taken these opportunities to educate site personnel that successive elevated viral loads in the absence of medications represent the potential for development of high-level resistance to medications. In addition, if a subject is not taking medications, then that regimen is not working, and has failed (especially when low-grade toxicities are present) for that individual.

As in any study, advances in treatment for HIV infection that occur after study initiation may affect the relevance of the study results. For instance, the development of much less toxic PI drugs might affect the relevance of a conclusion that PI-sparing regimens appear

better than NNRTI-sparing regimens because the PI drugs were more toxic. However, the validity of study results would not be affected, unless large numbers of subjects dropped out of the study in favor of these better alternatives.

It is never possible to demonstrate that the study results will be as relevant at the conclusion of any long-term study as at the outset, because the medical context in which the study takes place evolves during the course of the study. However, if we fail to conduct longer-term studies for fear that they may be irrelevant at their conclusion, we will never answer questions about long-term treatment strategies. Results from ACTG 384 are likely to remain relevant because the treatments in question have been demonstrated to have sufficient potency and tolerable toxicity and are therefore likely to be useful into the foreseeable future. Furthermore, the question of whether to initiate therapy with drugs from three rather than two classes is also likely to be of interest, even if new drugs in these classes are developed.

ACTG 384 is investigating longer-term treatment strategies, despite the challenges posed by recent, rapid advances in the clinical management of HIV. By using a factorial design that answers multiple questions and follows subjects through two or more sequences of regimens and choosing a primary study endpoint that reflects the multifactorial nature of treatment failure, the trial is intended to provide information about how initial choices affect subjects years after the start of therapy.

Acknowledgments

The authors wish to thank the entire ACTG 384 study team and the participants of ACTG 384. This work has been supported by grants AI-38855 and AI-38858 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

1. D'Aquila RT, Hughes MD, Johnson VA, et al. Nevirapine, zidovudine, and didanosine compared with zidovudine and didanosine in subjects with HIV-1 infection. *Ann Intern Med.* 1996; 124:1019–1030. [PubMed: 8633815]
2. Hammer SM, Squires KE, Hughes MD, et al. A controlled trial of two nucleoside analogues plus didanosine in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. *N Engl J Med.* 1997; 337:725–733. [PubMed: 9287227]
3. Gulick RM, Mellors JW, Havlir D, et al. Treatment with didanosine, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *N Engl J Med.* 1997; 337:734–739. [PubMed: 9287228]
4. Montaner JS, Reiss P, Cooper D, et al. A randomized, double-blind trial comparing combinations of zidovudine, didanosine, and zalcitabine for HIV-infected patients: The INCAS trial. *JAMA.* 1998; 279:930–937. [PubMed: 9544767]
5. Murphy RL, Gulick RM, DeGruttola V, et al. Treatment with didanosine alone or didanosine with zidovudine and lamivudine in adults with human immunodeficiency virus infection. *J Infect Dis.* 1999; 179:808–816. [PubMed: 10068575]
6. Friedland GH, Pollard R, Griffith B, et al. Efficacy and safety of didanosine mesylate with zidovudine and didanosine compared with two-drug combinations of these agents in persons with HIV disease with CD4 counts of 100 to 500 cells/mm³. *J Acquir Immune Defic Syndr.* 1999; 21:281–292. [PubMed: 10428106]
7. Staszewski S, Morales-Ramirez J, Tashima KT, et al. Efavirenz plus zidovudine and lamivudine, efavirenz plus didanosine, and didanosine plus zidovudine and lamivudine in the treatment of HIV-1 infection in adults. *N Engl J Med.* 1999; 341:1865–1873. [PubMed: 10601505]

8. Carpenter CC, Cooper DA, Fischl MA. Antiretroviral therapy in adults: Updated recommendations of the International AIDS Society. *JAMA*. 2000; 283:381–384. [PubMed: 10647802]
9. Deeks SG. Failure of HIV-1 protease inhibitors to fully suppress viral replication. Implications for salvage therapy. *Adv Exp Med Biol*. 1999; 458:175–182. [PubMed: 10549390]
10. Palmer S, Shafer RW, Merigan TC. Highly drug-resistant HIV-1 clinical isolates are cross-resistant to many antiretroviral compounds in current clinical development. *AIDS*. 1999; 13:661–667. [PubMed: 10397560]
11. Boden D, Markowitz M. Resistance to human immunodeficiency virus type 1 protease inhibitors. *Antimicrob Agents Chemother*. 1998; 42:2775–2783. [PubMed: 9797203]
12. Shafer RW, Winters MA, Palmer S, et al. Multiple concurrent reverse transcriptase and protease mutations and multidrug resistance of HIV-1 isolates from heavily treated patients. *Ann Intern Med*. 1998; 128:906–911. [PubMed: 9634429]
13. Izopet J, Bicart-See A, Pasquier C, et al. Mutations conferring resistance to zidovudine diminish the antiviral effect of stavudine plus didanosine. *J Med Virol*. 1999; 59:507–511. [PubMed: 10534734]
14. Kuritzkes DR, Quinn JB, Benoit SL, et al. Drug resistance and virological response in NUCA 3001, a randomized trial of lamivudine (3TC) versus zidovudine (ZDV) versus ZDV plus 3TC in previously untreated patients. *AIDS*. 1996; 10:975–981. [PubMed: 8853730]
15. Havlir DV, Richman DD. Viral dynamics of HIV: Implications for drug development and therapeutic strategies. *Ann Intern Med*. 1996; 124:984–994. [PubMed: 8624066]
16. Gulick RM, Mellors JW, Havlir D, et al. Simultaneous vs sequential initiation of therapy with indinavir, zidovudine, and lamivudine for HIV-1 infection: 100-week follow-up. *JAMA*. 1998; 280:35–41. [PubMed: 9660361]
17. Staszewski S, Miller V, Sabin C, et al. Rebound of HIV-1 viral load after suppression to very low levels. *AIDS*. 1998; 12:2360. [PubMed: 9863886]
18. Schapiro JM, Winters MA, Lawrence J, et al. Clinical cross-resistance between the HIV-1 protease inhibitors saquinavir and indinavir and correlations with genotypic mutations. *AIDS*. 1999; 13:359–365. [PubMed: 10199226]
19. Gilbert P, DeGruttola V, Hammer S. Efficient trial designs for studying combination antiretroviral treatments in patients with various resistance profiles. *J Infect Dis*. 1998; 178:340–348. [PubMed: 9697713]
20. DeGruttola V, Hughes M, Gilbert P, et al. Trial design in the era of highly effective antiviral drug combinations for HIV infection. *AIDS*. 1998; 12(Suppl 1A):S149–S156. [PubMed: 9632997]
21. Cox, DR. *Planning of Experiments*. John Wiley & Sons; New York: 1958.
22. Box, GEP.; Hunter, WG.; Hunter, JS. *Statistics for Experimenters*. Wiley & Sons; New York: 1978.
23. Williams, PL. Sample size calculations for failure time data. In: Finkelstein, DM.; Schoenfeld, DA., editors. *AIDS Clinical Trials*. Wiley-Liss; New York: 1995.
24. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *JASA*. 1999; 94:1096–1146.
25. Collett, D. *Modelling Survival Data in Medical Research*. Chapman & Hall; New York: 1994.
26. Mantel N. Ranking procedures for arbitrarily restricted observations. *Biometrics*. 1967; 23:65–78. [PubMed: 6050473]
27. Friedman, LM.; Furberg, CD.; DeMets, DL. *Fundamentals of Clinical Trials*. 3rd ed. Springer; New York: 1998.
28. Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons; New York: 1980.
29. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer*. 1976; 34:585–612. [PubMed: 795448]
30. DeGruttola V, Dix L, D'Aquila R, et al. The relation between baseline HIV drug resistance and response to antiretroviral therapy: Re-analysis of retrospective and prospective studies using a standardized data analysis plan. *Antiviral Therapy*. 2000; 5:41–48. [PubMed: 10846592]

Table 1

Study Schema

Arm	First Regimen	Second Regimen	Salvage
A	ddl + d4T + EFV	⇒ ZDV/3TC + NFV	⇒ ddl + hydroxyurea + IDV + APV ^a
B	ddl + d4T + NFV	⇒ ZDV/3TC + EFV	⇒ ddl + hydroxyurea + IDV + APV ^a
C	ZDV/3TC + EFV	⇒ ddl + d4T + NFV	⇒ ddl + hydroxyurea + IDV + APV ^a
D	ZDV/3TC + NFV	⇒ ddl + d4T + EFV	⇒ ddl + hydroxyurea + IDV + APV ^a
E	ddl + d4T + NFV + EFV	⇒ (directly to salvage)	⇒ ZDV/3TC + IDV + APV ^b
F	ZDV/3TC + NFV + EFV	⇒ (directly to salvage)	⇒ ddl + d4T + IDV + APV ^c

Double-blinded medications are in bold. ddl = didanosine; d4T = stavudine; EFV = efavirenz; ZDV/3TC = zidovudine plus lamivudine in a single pill; NFV = nelfinavir; IDV = indinavir; APV = amprenavir.

^aThis salvage regimen was changed (July 2000) to a choice of ritonavir + amprenavir (RTV + APV) or ritonavir + indinavir (RTV + IDV) with one of (ddl + hydroxyurea), (abacavir + ZDV/3TC), or (abacavir + d4T + 3TC).

^bThis salvage regimen was changed (July 2000) to a choice between ritonavir + amprenavir (RTV + APV) or ritonavir + indinavir (RTV + IDV), in combination with ZDV/3TC.

^cThis salvage regimen was changed (July 2000) to a choice between ritonavir + amprenavir (RTV + APV) or ritonavir + indinavir (RTV + IDV), in combination with ddl + d4T.

Table 2

Study Arms with First Treatment Combination (and Target Sample Sizes)

NRTI Factor	PI/NNRTI Factor		
	Initial NNRTI Efavirenz (EFV)	Initial PI Nelfinavir (NFV)	Initial PI + NNRTI NFV + EFV
ddI + d4T	A (125)	B (125)	E (150)
ZDV/3TC	C (125)	D (125)	F (150)

PI = protease inhibitor; NNRTI = non-nucleoside reverse transcriptase inhibitor; NRTI = nucleoside reverse transcriptase inhibitor; ddI = didanosine; d4T = stavudine; ZDV/3TC = zidovudine plus lamivudine in a single pill.

Table 3

First Objective—Sequences of PI, NNRTI versus Both Drug Classes Initially

NRTI Factor	PI/NNRTI Factor		
	Initial NNRTI	Initial PI	PI + NNRTI
ddI + d4T	A	B	E
ZDV/3TC	C	D	F
First Comparison:	↑	versus	↑
Second Comparison:		↑ versus	↑

PI = protease inhibitor; NNRTI = non-nucleoside reverse transcriptase inhibitor; NRTI = nucleoside reverse transcriptase inhibitor; ddI = didanosine; d4T = stavudine; ZDV/3TC = zidovudine plus lamivudine in a single pill.

Table 4

Second Objective—Sequence Starting with an NNRTI versus Sequence Starting with a PI

NRTI Factor	PI/NNRTI Factor	
	Initial NNRTI	Initial PI
ddI + d4T	A	B
ZDV/3TC	C	D
Comparison:	↑	↑

PI = protease inhibitor; NNRTI = non-nucleoside reverse transcriptase inhibitor; NRTI = nucleoside reverse transcriptase inhibitor; ddI = didanosine; d4T = stavudine; ZDV/3TC = zidovudine plus lamivudine in a single pill.

Table 5

Third Objective—ddI + d4T Followed by ZDV/3TC versus the Converse Sequence of NRTIs

NRTI Factor	PI/NNRTI Factor		Comparison
	Initial NNRTI	Initial PI	
ddI + d4T	A	B	⇐
ZDV/3TC	C	D	⇐

PI = protease inhibitor; NNRTI = non-nucleoside reverse transcriptase inhibitor; NRTI = nucleoside reverse transcriptase inhibitor; ddI = didanosine; d4T = stavudine; ZDV/3TC = zidovudine plus lamivudine in a single pill.

Table 6

Baseline Characteristics

	Total (n = 980)
Gender	
Male	799 (82%)
Female	181 (18%)
Race/ethnicity	
White	456 (47%)
Black	339 (35%)
Hispanic	163 (17%)
Asian	17 (2%)
American Indian	4 (0%)
Missing	1 (0%)
IV drug use	
Never	890 (91%)
Currently	8 (1%)
Previously	82 (8%)
Age (years)	
Median	36.0
13–29	214 (22%)
30–39	418 (43%)
40–49	244 (25%)
50–59	78 (8%)
60 and over	26 (3%)
CD4 count (cells/mm ³)	
Median	277.5
<100	248 (25%)
100–199	115 (12%)
200–299	162 (17%)
300–500	275 (28%)
>500	179 (18%)
Missing	1 (0%)
Plasma HIV–1 RNA copies/mL	
Median (log ₁₀ copies/mL)	4.94
<50	1 (0%)
50–499	10 (1%)
500–4999	73 (7%)
5000–49,999	293 (30%)
50,000–499,999	428 (44%)
≥500,000	175 (18%)