DRU-1333-P

March 1996

RAND

The Medicine Education and Research Foundation, San Diego

Measuring Patient-Centered Outcomes in the Pfizer-CCTG Trial of Azithromycin vs. Rifabutin vs. Both for the Prevention of HIV-Associated Mycobacterium avium: Instruments, Methods, Application, and Results

Prepared for Pfizer by
Samuel A. Bozzette
Sandra H. Berry
Brian C. Wright
David E. Kanouse

With the assistance of
Naihua Duan
Carol Edwards

PREFACE

This report summarizes results of work sponsored by Pfizer and aimed at capturing the overall outcomes of *Mycobacterium avium* prophylaxis in persons with advanced HIV disease. While this report concentrates on the Pfizer/California Collaborative Treatment Group (CCTG) trial comparing azithromycin to rifabutin and to the combination of both drugs, it also reviews selected results from an entire program of research to develop instruments and analytic methods for capturing health status and other patient-centered outcomes in AIDS clinical trials. We describe this broader context to provide additional information on the provenance of these instruments and procedures, which now form the basis for current standard approaches to capturing these outcomes in HIV clinical trials.

Inasmuch as fungal endpoints data were not available to us at the time the *Mycobacterium avium* results were being developed, a supplemental report covering these endpoints will be issued at a later date.

CHAPTER I
BACKGROUND

Acquired immunodeficiency syndrome (AIDS) is a disease characterized by profound immunosuppression that results in increased susceptibility to a variety of opportunistic infections and the occurrence of unusual forms of neoplasms. It is currently estimated that up to 1 million persons in the United States have been infected with HIV, and it is expected that a large proportion of these individuals will develop AIDS. The vast majority of clinical research studies on HIV infection and AIDS have focused on clinical endpoints, such as time to opportunistic infection or death, to determine therapeutic efficacy of investigational agents.

Because of the emphasis on improving survival or preventing opportunistic infection, less attention has been paid to measuring the therapeutic impact of new agents on general health status or on health-related quality of life. In clinical practice, however, the choice between alternative therapies often rests not on their prophylactic efficacy or effects on survival, which may be quite similar, but on differences in symptom relief and side effects--that is, on which agent can be expected to yield the best level of functioning and quality of life for the patient over the course of treatment. Even where agents have different effects on survival, health status and quality of life are of concern to physicians and patients; some patients may prefer to live (or remain free of opportunistic infection) a shorter time with higher quality of life rather than have survival (or freedom from opportunistic infection) prolonged by a therapy that produces life-interfering side effects and adverse reactions.

For these reasons, assessing health status and quality-of-life may add valuable information to clinical trials that focus largely upon survival status or freedom from opportunistic infections as final endpoints, potentially broadening the basis on which new agents are evaluated for approval and informing subsequent treatment decisions by physicians and patients. To allow for such evaluation, it is important to develop reliable and valid measures that are sensitive to changes in the various domains that health status and quality of life encompass--e.g., physical functioning, distress from physical symptoms, emotional well-being, cognitive functioning, social functioning, role functioning, and general health perceptions. In addition to being reliable and valid, such measures need to be

suitable for administration to a clinical population in the field settings represented in a clinical trial.

With this perspective in mind, staff at RAND and University of California at San Diego undertook a systematic assessment of health status and quality of life outcomes as part of ACTG 081, a clinical trial of prophylaxis against HIV-related Pneumocystosis and invasive mycoses, and ACTG 981, a nested substudy of preventive therapy for HIV-related fungal disease. Soon after that trial began, parallel studies were initiated for ACTG 114, a trial of zalcytabine versus zidovudine, ACTG 116/117, a clinical trial of didanosine at two different doses versus zidovudine in patients with advanced HIV disease. About the same time, we undertook for Pfizer an assessment of health status outcomes of prophylaxis for *Mycobacterium avium* bacteremia. This report summarizes the developmental work performed for these trials, the field experience administering the measures, and the findings that emerged from analyses of health status and quality-of-life outcomes for the Pfizer/CCTG prophylaxis trial.

MEASUREMENT AND ANALYTIC APPROACH

Drug therapy for HIV-related illness is prescribed to improve a broad range of patient health outcomes, including survival and various medical components of health status. The health outcomes of concern in prescribing a particular drug are often multidimensional, so that arriving at an overall assessment of the effects of treatment requires integrating information on various aspects of health status. This means weighing in some fashion, whether implicitly or explicitly, all relevant aspects of health status potentially affected by the drug, including both toxic and salutary effects. Quality-of-life and health status measures offer an attractive way to do this, in that they permit the effects of disparate health-states to be compared in a common metric based on how they affect the patients' experiences of their current health and satisfaction with it. In principle, this facilitates the process of making rational choices rooted in patient preferences among therapies that present qualitatively different risk/benefit profiles.

Although interest in patient-reported health status measures as outcomes in clinical trials has been growing, such measures have rarely been used as primary or secondary endpoints in actual trials. Among the reasons for this limited acceptance is the failure of common

analytic procedures to meet key requirements: (1) results of clinical trials should be expressed in terms that clinicians, regulators, administrators, and patients find useful; (2) reasonable means of summarizing outcomes should be available *a priori*; (3) scale units should have real-world meaning; (4) unwarranted assumptions regarding scale properties should be avoided; and (5) adequate methods for handling attrition from death or drop-out must be found.

Much of the work performed in connection with this trial and the ACTG 081/981 and ACTG 116/117 trials was to develop approaches for dealing with these problems. To address requirements (1) and (2), we adapted several scales from the Medical Outcomes Study (MOS) to provide the detail necessary to characterize the varied dimensions of health-related quality of life, but we created an overall summary score (the Perceived Health Index) to use in determining the best overall treatment across all dimensions. We used the highly reliable 5-item MOS Current Health Perceptions scale as a reference variable (Stewart, Hays and Ware, 1992), deriving weights that represent the best linear combination of domain-specific MOS health status scales for predicting Current Health Perceptions.

To address the need for scale units with real-world meaning (requirement 3), we extended the language and logic of survival analysis (and related concepts such as "disease-free survival") to describe health-related quality of life more generally. The effects of alternative treatments are compared in terms of the time that patients who receive them experience health states of at least a given quality. This approach shifts the comparison from units that lack intuitive meaning (such as average scores on an index of health-related quality of life) to units that are better understood (time). This approach also avoids the problem of unwarranted assumptions about scale quality (requirement 4): Because we do not need to average scores on health-related quality of life, our only requirement is that each scale value can be ranked relative to the others (ordinality assumption). It is much easier to order health states (and their corresponding scale values) than it is to assign an absolute value to each.

Once we have measured health states in a way that summarizes information across several domains and produces a score on an ordinal scale describing patients' overall health state,

we can employ a set of analytic techniques known as multistate survival analysis to draw proper inferences from repeated outcome measures in clinical trials. Multistate survival analysis, a generalization of survival analysis developed for use in this and other ACTG clinical trials, allows for the combination of data on survival and health status and provides reasonable ways to handle attrition, thereby addressing requirement (5) above. Multistate survival analysis is described in more detail below. First, however, we summarize the psychometric results for the Perceived Health Index.

Psychometric Results. The psychometric properties of the health status scales were assessed using multitrait scaling and test-retest stability. Weights for the index were derived from regressions of Current Health Perceptions on the domain-specific health status scales. The effect of participant characteristics on weights was tested with additional regressions and sensitivity analysis. Finally, the reliability and known-clinical-groups validity of the index were assessed.

Data were obtained from 1,862 participants in various randomized controlled clinical trials of chronic therapies for advanced HIV disease conducted by the AIDS Clinical Trials Group (ACTG) who provided a total of 7,352 observations (Bozzette, Hays, Berry, Kanouse, 1994). Over 50 percent of participants were from sites enrolling at least 96 percent of candidates. The mean CD4+ count was 131. The internal consistency reliability (Cronbach's alpha) of the multi-item scales ranged from 0.86 to 0.90, and items demonstrated excellent discrimination across scales. The domain-specific scales explained 59 percent of the variation in the Current Health Perceptions scale ($p < 0.00001$). The resulting Perceived Health Index was equal to 0.20*Physical Functioning + 0.15*Pain +0.41*Energy/Fatigue + 0.10*Emotional Well-Being + 0.05*Social Functioning +0.09*Role Functioning. A strong positive bivariate relationship between the Cognitive Function/Distress scale and the Current Health Perceptions scale was subsumed by the combination of the other domain-specific scales in multiple regressions, so it does not appear independently in the index. The proportional weights used in the index were insensitive to variations in demographics. The reliability of the index was conservatively estimated to be 0.94. Patients with index scores in the lowest quartile had a 2- to 11-fold higher probability than those in the highest quartile of reporting various specific clinical events, and the index correlated significantly more highly with the number of such events than did the Current Health Perceptions scale.

The modified health status scales included in the HIV-PARSE are reliable and valid in patients with advanced HIV disease. The Perceived Health Index provides a reliable and valid means of summarizing self-reported current health, correlates strongly with clinical indicators, and should be useful as an outcome measure in patients enrolling into clinical trials of therapies for advanced HIV disease.

Multistate Survival Analysis. The central notion of multistate survival analysis is that any observed scale score may be used as a threshold dividing participants into two groups, with those having higher-than-threshold scale scores being considered to have better health. For any threshold score, one calculates the total amount of time during the study that a subject has better-than-threshold health, or time-above-threshold state (TATS). Just as the usual survival curve is the complement of the survival time distribution, one can take the complement of the TATS distribution to obtain the survival above threshold state (SATS) curve. The SATS curve, a generalization of the usual survival curve, gives the proportion of subjects whose TATS for a given threshold is greater than each given total duration time. If the threshold score being considered is below the lowest observed score, all surviving subjects have better than threshold scores; in this situation, the TATS is equivalent to the usual survival time, and the SATS curve coincides with the usual survival curve. Thus, incorporating mortality into the analysis requires only that death be considered worse than all recorded scores.[1]

If the threshold score is higher than the lowest observed score, the SATS score is analogous to a "survival curve" for time above a clinically defined threshold, such as the proportion of patients not experiencing an opportunistic infection. However, the two curves are generally not identical because the standard survival curve depicts the chronological time until a one-time-only event, whereas the SATS curve depicts the total cumulative time free of the event (or above the threshold). Thus, the SATS curve captures an unlimited number of deteriorations and improvements over the course of the study. This contrasts with standard survival analysis, which can be used to describe the time until the first opportunistic infection or the first drop below a health status threshold but cannot address

[1] Allowing live states worse than death in an intervention study is inappropriate because this could allow a fatal side effect of treatment to improve overall health status.

the total duration of time free from opportunistic infections or with above-threshold health status.

In the calculation of the TATS and SATS, attrition is handled by extending the standard Kaplan-Meier assumption that attrition is uninformative regarding unobserved states. The TATS and SATS can be estimated for a large number of thresholds covering the entire range of observed health status scores. Averaging the TATS across both patients and thresholds gives a summary of the overall time/quality experience of the cohort; namely, the typical time that a typical patient in the cohort spends above a typical threshold. Plotting the SATS for a given duration of time against thresholds covering the range of observed scores yields a graphical summary of that experience, known as a SATS map.

DEVELOPMENT OF MEASURES

The following is a brief summary of the work carried out to develop the instrument and scales used in the Pfizer/CCTG trial. This work was supported in part by a grant from Pfizer to UCSD/RAND through the Medical Education and Research Foundation. Later chapters of this report and appended publications provide more detail.

Instrument Development. We developed the HIV-PARSE instrument specifically to measure patient-reported global health status and functioning, symptom impact, disability, work, and health care service utilization of patients with HIV disease. Chapter 2 describes the instrument, how it was developed, the modifications made for Pfizer/CCTG Study 174, and how it has been changed as a result of experience in this trial and the ACTG trials.

Administration of the HIV-PARSE Instrument. The HIV-PARSE instrument was administered to men and women enrolled across 12 treatment sites in Pfizer/CCTG trial for the prevention of *Mycobacterium avium* bacteremia. Questionnaires were administered in clinic settings, at the time of clinic visits; patients generally filled them out in the waiting area. We sought to obtain completed instruments at baseline and at 16, 32, 48, 72, and 96 weeks thereafter.

The questionnaires were distributed and collected by personnel at the CCTG study sites according to a protocol developed by RAND, and transmitted to Pfizer via usual channels for data entry. Our field experience is summarized in Chapter 2.

Scale Development. As described above, we developed a Perceived Health Index that combines measures of specific health domains considered relevant to treatment into an overall summary of health status. This work is described in Bozzette, Hays, Berry, and Kanouse (1994).

Analytic Methods. Some of the problems that one encounters in attempting to include quality-of-life outcomes in clinical trials and our conceptual approach to dealing with these problems are described in Bozzette, Duan, Berry, and Kanouse (1994). The analytic methods are described in more detail in the following papers that have been submitted for publication:

> Bozzette SA, Duan N, Kanouse DE. Multistate survival analysis I: Time above threshold state.
>
> Bozzette SA, Duan N, Kanouse DE. Multistate survival analysis II: Generalized Mantel-Haenszel tests of transitions from states.

APPLICATIONS IN OTHER CLINICAL TRIALS

The methods described above were applied to analyze health status and quality-of-life outcomes in ACTG 081, a trial of prophylaxis against first episode Pneumocystosis, and, ACTG 981, a trial of prophylaxis against invasive mycoses, in HIV-infected persons at least 13 years of age with a history of having a CD4+ lymphocyte count below 200 per milliliter. In ACTG 081, all three drugs (trimethoprim-sulfamethoxazole, dapsone, and aerosolized pentamidine) were known to be active but to have differing toxicity profiles, which could make it difficult to arrive at overall conclusions regarding therapeutic effects based on traditional clinical endpoints. In ACTG 981, only those randomized to fluconazole rather than clotrimazole received an active agent, but the relatively low incidence of the invasive mycoses raised questions regarding the overall benefit, even if prophylaxis was effective.

Clinical data indicated that differences in prophylactic efficacy between systemic and aerosolized pentamidine therapy were not large. Analysis of data from 785 patients who completed a baseline HIV-PARSE form and at least one follow-up showed that mean scores on all MOS scales were essentially identical for all scales and indices except for the Current Health Perceptions scale, which was 0.15 standard deviations higher in the aerosolized pentamidine group. Use of hospital and other medical services was similar across the three treatment groups, except for a slightly higher use of non-protocol medications in the aerosolized pentamidine group. Trimethoprim/sulfamethoxazole recipients reported slightly fewer days in bed, days of reduced activity, and days feeling less well than usual. Multistate survival analysis on the Perceived Health Index revealed that typical entrants in the trimethoprim/sulfamethoxazole, dapsone, and aerosol pentamidine arms spent 12.9, 12.5, and 13.1 months, respectively, in at least a typical health state ($p = 0.27$ to 0.66).

When subgroups with fewer than 100 CD4+ cells were analyzed separately, functional and health status measures became less favorable, and differences between treatment groups were larger and more consistent. On essentially all these measures, dapsone recipients reported more favorable outcomes. For the Perceived Health Index and the Mental Health/Emotional Well-Being scale, the difference exceeded 0.20 standard deviations. Dapsone recipients also reported fewer symptoms and fewer symptoms that interfered with their functioning. Recipients of systemic therapy reported more days in bed, more days of reduced activity, and more days feeling less well than usual compared with those in the aerosolized pentamidine arm, with differences ranging from 5 to 12 percent. Differences in health care use were small and non-significant, except for greater use of non-protocol medications by the aerosolized pentamidine group. In this same subgroup, multistate survival analysis on the Perceived Health Index showed that typical entrants in the trimethoprim/sulfamethoxazole, dapsone, and aerosolized pentamidine arms spent 12.4, 12.5, and 11.3 months, respectively, in at least a typical health state ($p = 0.30$ to 0.95). However, analysis of mean time above threshold state (MTATS) suggested a trend toward better-quality survival among those randomized to systemic therapy.

In ACTG 981, outcomes differed very little overall between the two arms. Statistically significant differences were found favoring clotrimazole on the Cognitive Functioning/

Distress scale and fluconazole on the Pain scale, but only 3 of 29 non-independent comparisons demonstrated statistically significant differences. Multistate survival analysis on the Perceived Health Index revealed that typical entrants in both the clotrimazole and fluconazole arms spent 9.7 months in at least the typical health state, and MTATS data indicated that quality-survival outcomes for the two treatment groups were very similar.

In the subgroup of patients entering with fewer than 50 CD4+ cells, outcomes captured by the MOS scales were again similar for the two treatment groups, except that scores on the Current Health Perceptions scale were higher in the fluconazole group. However, both the hospitalization rate and the average number of hospital days were approximately 41 percent higher in the clotrimazole group ($p = .03$ and .01, respectively). Recipients of clotrimazole reported more symptoms, used more procedures, home care, and medications and had more telephone contacts with providers. They reported between 3 percent and 10 percent more days in bed, days of reduced activity, and days of missed work. Multistate survival analysis on the Perceived Health Index showed that a typical person entering 981 with fewer than 50 CD4+ cells at baseline spent 8.6 months with at least the typical health state if assigned to clotrimazole and 9.4 months if assigned to fluconazole ($p = 0.09$).

HIV-PARSE measures and the analytic methods described above were also applied in a study of the outcomes of prescribing zalcitabine (ddC) or zidovudine (AZT) for initial therapy of advanced HIV disease (Bozzette, Kanouse, Berry, and Duan 1995). Patients participating in this trial had HIV infection, fewer than 200 CD4+ cells, and either a history of *Pneumocystis carinii* or symptoms of HIV infection. This substudy included 58 percent (338/668) of main study enrollees representing 90 percent of enrollees at participating sites. They were prescribed either zalcitabine at 0.75 mg every eight hours plus inactive capsules identical in appearance to zidovudine or zidovudine at 200 (later 100 mg) every four hours plus inactive tablets identical in appearance to zalcitabine. Outcomes were assessed by a self-report survey instrument containing specific questions about disability, work, functioning, and utilization as well as nine health and functioning scales adapted from the MOS.

Differences between the treatments were striking. Zalcitabine recipients were twice as likely to undergo an invasive procedure ($p = .004$) or be admitted to a hospital ($p = 0.01$).

Zalcitabine recipients reported greater than 40 percent more symptoms that interfered with their activity ($p = .001$) and greater than 50 percent more disability days ($p < .01$). They also had a 7 percent lower employment rate and a 35 percent lower monthly income. Average observed health status scores were lower in zalcitabine recipients overall, but especially in the early portion of the study. When survival and health-status data were combined using multistate survival analysis, results showed that, over 76 weeks of study, a typical zidovudine recipient spent about four (10 percent) more weeks with at least the typical health state than did a typical zalcitabine recipient.

The methods described above were also applied to analyze health status and quality-of-life outcomes in a study comparing the effects of zidovudine and didanosine in persons with advanced HIV infection (Bozzette, Kanouse, Duan, Berry, and Richman, 1996). Patients were 356 participants enrolled in ACTG 116 and 117. All had HIV infection and either a CD4+ cell count < 200 or a CD4+ cell count < 300 plus symptoms of HIV disease. Participants were randomized equally within strata defined by duration of prior zidovudine therapy, to receive didanosine sachets at a dose of 500 mg daily (334 mg in subjects weighing < 60 mg) or 750 mg daily (500 mg in subjects weighing <60 kg) plus inactive capsules resembling zidovudine, or to receive zidovudine capsules at a dose of 600 mg daily plus inactive sachets resembling didanosine.

Data from the HIV-PARSE instrument showed no differences in reported symptom impact or health care utilization, and most measures of disability were similar. In the group with more than eight weeks of prior zidovudine therapy, several of the health status scale scores for ongoing participants were significantly better for didanosine recipients, but average differences were small. Use of several approaches to combining health status and survival showed no differences in the overall quality/time experiences between the treatment groups. Individuals taking zidovudine, low-dose didanosine, and high-dose didanosine had 33, 34, and 35 weeks, respectively, in at least the typical health state if they had fewer than eight weeks of previous zidovudine, and had 23, 23, and 26 weeks, respectively, if they had more than eight weeks previous use of zidovudine. Results did not differ when data were analyzed within strata of persons who had any versus no prior exposure to zidovudine, or AIDS versus non-AIDS status. Functional status and health-related quality of life were largely similar among persons receiving either zidovudine or didanosine,

regardless of the duration of prior zidovudine treatment. Thus, the treatment differences demonstrated in the clinical analyses of these trials (Kahn et al., 1992) did not translate into substantial differences in functional status or health-related quality of life for typical patients in the substudy.

CHAPTER II: THE HIV-PARSE INSTRUMENT: DEVELOPMENT AND FIELD EXPERIENCE

DEVELOPMENT OF THE HIV-PARSE INSTRUMENT

As noted in Chapter I, measurement of health status and functioning may be a useful adjunct to traditional outcome measures, providing a direct measure of health in the absence of marker clinical events and an integrated measure of overall effect in patients who do and do not experience such events.

ACTG 081 represented the first of the NIAID HIV-related clinical trials to include health status as an official outcome of the trial. The HIV-PARSE instrument, which is available as RAND Publication MR-342-NIAID was developed to measure patient reports of health status (Berry et al., 1994). This instrument was revised for the Pfizer-CCTG *Mycobacterium avium* prophylaxis trial.

OVERVIEW OF QUESTIONNAIRE

The HIV-PARSE instrument is designed in four main sections: demographic background, risk group and life circumstances; health status, including a quality of life measure; utilization of health services; and a checklist of symptoms and symptom impact. The purpose and source of items for each section are described below. The scoring rules for the scales are explained and the reliability of the scales, based on an HIV clinical trial population, are provided. Appendix A of MR-342-NIAID is the original version of the form used in the ACTG 081 trial. The source listing for each item in the Appendix A questionnaire is included as Appendix B of the MR. Appendix C of the MR is the version of the instrument used in the Pfizer-CCTG trial in a combined English-Spanish version.

Background, Risk Group, and Life Circumstances

This section includes items on highest year of schooling, job history during the past year and the past week, type of work respondent does, total personal income for the past year and the past month, and type of health insurance coverage. The main purpose of this section is to provide explanatory or control variables to use in the analysis of the effects of

the drug on patient functioning and utilization of health services. In addition, certain variables were included to serve as outcome variables for some longitudinal analyses; for example, change in work status can be tracked over time.

Most items in this section were drawn from standard sources such as the MOS, the US Census, and the NORC General Social Survey, often with some modification for this survey. For example, the type-of-work item is drawn from the US Census, with examples associated with each category selected to include type of jobs common to the population of the study. The type-of-insurance item includes a category indicating whether the respondent uses his usual insurance to pay for HIV care (some HIV patients avoid doing this to protect their privacy).

Health Status and Quality of Life

The second section is mainly drawn from items in the 24-month Patient Assessment Questionnaire, a self-administered questionnaire used in the MOS and administered by mail two years after entry into the study.[2] Modifications were made, however, to improve the quality of measurement for an HIV population versus a general patient population.[3] Measures in this section include:

Current health perceptions
Physical functioning
Pain
Energy/Fatigue
Emotional well-being
Psychological distress and well-being
Cognitive functioning
Social functioning
Role functioning
Quality of life
Will to function
Disability days

[2]A full description of the MOS measures can be found in Stewart and Ware, 1992.
[3]The main modification was to change the wording of the role functioning response categories from "yes, for more than 3 months," "yes, for less than 3 months," or "no" to "yes, all of the time," "yes, some of the time," or "none of the time."

These items are designed to serve as outcome measures that tap specific dimensions of self-reported health and quality of life. The reference period for these items is the implicit "now" for current health perceptions, physical functioning, quality of life, and will to function. For all other measures, it is the four-week period prior to administration of the questionnaire. Scoring rules for the key scales are shown in Tables 2.4 and 2.5. Scale means and standard deviations are shown in Table 2.3 in this section.

Utilization of Health Services

The next section of the instrument measures use of health services, including whether the respondent had seen various types of providers and the number of contacts with each type of provider in the past four weeks, the number of home visits from doctors, nurses or other health professionals, the number of telephone contacts with all providers, the number of overnight hospital stays, how many stays, and how many nights in the hospital and in ICU. Again, the reference period in this section is the implicit "now" or the four weeks prior to administration of the questionnaire, as described above. Patients have been shown to be reliable reporters of medical care process events (Brown and Adams, 1992). The questionnaire also asks about use of paid personal care assistance and the number of hours used in the past four weeks. These items were designed to allow comparison of effects of drug combinations involved in the trial through comparison of specific items (e.g., treatments, diagnostic procedures), counts of items, or dollar values of item bundles, based on a value scale for the items.[4] This scale is currently being selected and tested. Most of these items were adapted from the MOS.

Symptom and Symptom Impact

This section asks whether the respondent has experienced each of 22 symptoms or groups of symptoms during the previous four weeks. The respondents were asked to rate the impact of each symptom experienced in terms of how much it interfered with normal

[4]A value scale provides a standard dollar value associated with a particular treatment or procedure, based on an average cost or price. It is used to compare the economic value of services or treatments provided in different care systems, where cost accounting procedures may result in differences in observed price or where dollar values may not be provided (e.g., within a health maintenance organization where care is prepaid).

activities. The symptom list was drawn from existing items, augmented to reflect clinical experience treating patients with advanced HIV disease as reported in the ACTG database. The symptom list was used as a direct measure of the impact of side effects for alternative treatment regimens as well as to facilitate analyses of the differential impact of various symptoms on health status and quality of life.

PRETESTING OF THE HIV-PARSE QUESTIONNAIRE

The questionnaire was originally pretested with approximately 40 adult patients at the UCSD AIDS Clinical Trials Group (ACTG). Patients completed the questionnaire, recording the time at which they began and ended. In addition, cognitive pretesting was carried out--ten patients completed the questionnaire in the presence of an interviewer who timed the administration and answered any urgent questions. After completion of the form, the interviewer debriefed each pretest respondent to determine whether any portions of the questionnaire were unclear, confusing or offensive. This included not only asking the respondent to raise problems, but also probing for understanding of key concepts in the questionnaire and for strategies used by respondents to answer the questions. Pretest instruments were data entered and checked for scale-properties (Cronbach's alpha reliability) and distributions of key variables (checking to be sure response categories captured variation among the respondents).

MODIFICATIONS TO THE HIV-PARSE

Several modifications were made to the form in later versions of the instrument. The changes are noted briefly in Table 2.1, along with the reasons for these changes. Items deleted to reduce length were simply combined or dropped to reduce the administration time and respondent burden. In a study with these topics as a central focus, the longer version could be used. In addition, the questionnaire has been translated into Spanish. Questions were first translated from English to Spanish, then back-translated from Spanish to English, with discrepancies resolved by discussion among the translators and the investigators. The Spanish version was administered in several locations, including Florida, New York, and Southern California, and the wording did not pose problems in the field. However, the psychometric properties of the Spanish version have not been

completely evaluated. The version of the HIV-PARSE questionnaire used in the Pfizer-CCTG trial (English/Spanish version) is shown in Appendix C to MR-341-NIAID for the convenience of readers.

The HIV-PARSE instrument has been administered to at least 675 of the 723 participants in the Pfizer-CCTG trial of azithromycin for the prevention of *Mycobacterium avium* across 12 treatment sites. The sites were instructed to administer the questionnaire in clinic settings, at the time of clinic visits. The questionnaire is provided to the patient for self-administration, usually in the waiting area. The questionnaires were to be administered at baseline and four-month intervals for the first 48 weeks of follow-up and at six-month intervals for the second 48 weeks (that is, at study weeks 0, 16, 32, 48, 72, and 96).

Table 2.2 shows the number of registrants in the clinical protocol for each site and the number of participants completing the HIV-PARSE portion of the registration process. The last column shows the percentage of clinical protocol enrollees in each site who also took part in the HIV-PARSE protocol. The proportions ranged from 80 to 99 percent across sites, with an overall mean participation rate of 93 percent.

To evaluate nonresponse, we define eligible respondents as those registrants who went on to qualify for the study, enroll, be randomized, and ultimately be evaluated. In this study, 29 registrants did not complete the baseline evaluation, and 19 were not randomized, presumably because they were demonstrated not to qualify for the study. Of the remaining 675 participants, 26 were dropped from the analytic dataset by the sponsor because they were not demonstrated to have negative cultures for *Mycobacterium avium* at baseline. We define the response rate as the number of questionnaires received divided by the number of questionnaires expected from 649 patients who could be clinically evaluated at the designated survey weeks, taking into account reported discontinuations. This definition leads us to discard relatively few surveys, as most ineligible individuals completed only one or two. Overall, 2,065 of the total 2,113 responses received were from the 649 participants who could be evaluated.

Table 2.1

CHANGES MADE TO HIV-PARSE INSTRUMENT

TYPE OF CHANGE	REASON FOR CHANGE
Dropped demographic items	Captured in other parts of study protocol.
Reduced the section on work history in the previous 12 months from four to two items.	Reduce length
Dropped the item on importance of job skills	Reduce length
Added a visual scale quality-of-life item	Additional summary measure, anchors to death as a state
Dropped items on living situation	Reduce length
Added new items on social functioning and pain	Replaced single item with multiple items to improve precision of measurement
Dropped items on emotional and behavioral control	Did not add unique information over and above other emotional well-being items in terms of predicting overall health perceptions
Reduced number of types of providers listed and added number of visits for each provider type, omitted asking about telephone contacts for each specific provider type	Reduce length and improve specificity of information
Added items on group home or hospice stays and paid in-home assistance with bathing, dressing, or household chores	New items to measure important types of utilization
Reduced symptom checklist to 22 items by combining items	Reduce length
Added item on assistance with completing the questionnaire	Methodological interest in assessing respondent and staff burden

Table 2.2

AZITHROMYCIN ENROLLMENT BY SITE

SITE	CLINICAL	PARSE	%
26B	38	34	89
30A	98	91	93
32A	59	53	90
41A	35	28	80
57B	86	85	99
66A	23	22	96
75B	78	73	94
842	67	63	94
935	86	84	98
94B	21	18	86
968	16	15	94
981	116	109	94
TOTAL	723	675	93

SOURCE: Calculated from data supplied by Pfizer.

FIELD EXPERIENCE

In Table 2.3, these 2,065 questionnaires received are broken down into the baseline and scheduled follow-up waves. Within wave, the table indicates how many questionnaires were completed within one week of the scheduled date (on schedule), more than one week after the scheduled date (off schedule), and during clinic visits occurring between scheduled survey waves (interim). The percentage of active enrollees who completed forms for the baseline wave was 97 percent. For follow-up study weeks, completion rates varied from 91 percent (632/649) at the first follow-up to 66 percent (51/77) after 2 years and 10 percent (1/10) at week 112; however, rates were at least 79 percent through the first year of the clinical study, when the vast majority of clinical data were collected. Overall, the total of 2,065 forms represented 87 percent of expected forms plus 128 interim forms. Of the total forms received, 61 percent were completed within a week of the scheduled visit, another 33 percent were completed more than one week from the scheduled visit date, and 6 percent were completed in interim weeks.

Table 2.3

PERCENTAGE OF PARSE FORMS RECEIVED BY STUDY WEEK

Study Week	Number of Enrolled Patients	Number of Forms Received on Schedule	Number of Forms Received off Schedule	Total Number of Scheduled Forms Received	Percentage of Expected Forms Received	Number of Interim Forms Received	Total Number of Forms Received
0	649	611	21	632	97%	0	632
16	525	319	161	480	91%	16	496
32	430	181	187	368	86%	14	382
48	348	96	178	274	79%	13	287
72	185	48	83	131	71%	41	172
96	77	11	40	51	66%	29	80
97-112						16	16
TOTAL	2214	1266	670	1936	87%	129	2065

As indicated elsewhere, large attrition after a diagnosis of *Mycobacterium avium* disease is a problem with this dataset. As a consequence, certain analyses are best done using only data obtained prior to an endpoint diagnosis. Table 2.4 indicates the response rate by week for participants prior to any diagnosis of *Mycobacterium avium* disease. The difference between Table 2.3 and Table 2.4 is that 36 data points from 23 patients are dropped. With this restriction, the total number of eligible forms received drops to 2,029, but the overall completion rate climbs to 87 percent of expected scheduled forms and 93 percent overall.

Finally, it is important to note that the total of 2,065 forms received represents the largest potential sample size for analysis, but that this potential could not be fully used. Analysis for differential treatment effects required that persons complete a baseline form and at least one follow-up form; only 502 of the 611 enrollees completing a baseline form also completed follow-up forms, and some patients completing follow-up forms did not complete baselines. As a result, the main analyses used a total of 1,894 responses.

Table 2.4

PERCENTAGE OF PRE-*MYCOBACTERIUM AVIUM* PARSE FORMS RECEIVED BY STUDY WEEK

Study Week	Number of Enrolled Patients	Number of Forms Received on Schedule	Number of Forms Received off Schedule	Total Number of Scheduled Forms Received	Percentage of Expected Forms Received	Number of Interim Forms Received	Total Number of Forms Received
0	649	611	21	632	97%		632
16	520	316	157	473	91%	14	487
32	426	179	184	363	85%	13	376
48	344	93	174	267	78%	12	279
72	178	47	78	125	70%	37	162
96	74	11	38	49	66%	28	77
97-112						16	16
Total	2191	1257	652	1909	87%	120	2029

SPECIAL STUDY OF PATIENT AND STAFF REACTIONS TO ADMINISTRATION OF THE SURVEY

To address concerns about burden and confidentiality, we conducted a special survey of 94 patients and 20 members of the clinic staffs in March, 1991 at six ACTG sites. On this survey, respondents reported that the 081 baseline form took a mean of 18.7 minutes to complete. (The follow-up version is considerably shorter than the baseline instrument because it contains 18 fewer items.) The revised baseline questionnaire takes 10 to 15 minutes to complete. Most patients reported that they completed the HIV-PARSE baseline form in the waiting area; about a third finished it while waiting in the examination room. Only 5 percent required assistance with the questionnaire or used audio tapes provided for people with vision or reading problems, although 5 percent said they had some difficulty reading the questionnaire and 7 percent reported difficulty with writing responses.

Confidentiality concerns arose about the role of clinic staff in handing out and collecting the questionnaires and in copying and filing the completed instruments (according to NIAID procedures, a copy of all forms must be kept on site). In the special survey, most patients thought that their HIV-PARSE questionnaire responses were routinely seen by clinic staff (only 9 percent thought staff did not look at responses), but no one felt that having staff see

responses was "a bad thing." In fact, 37 percent thought it was "a good thing," 44 percent were indifferent, and 19 percent were "not sure." (The special survey was sealed in an envelope and mailed back directly; it was *not* seen by clinic staff.)

AIDS is a very difficult disease, and the PARSE instrument requires respondents to focus on its effects in a very systematic and personal way. Patients reported mixed views of their experience in filling out the form over the course of many visits. We asked patients to agree or disagree[5] with a series of statements about the PARSE form. The results are shown in Table 2.5.

Table 2.5

PATIENTS' VIEWS OF HIV-PARSE INSTRUMENT

Percentage Who Indicated They "Agreed Strongly" or "Agreed Somewhat" with Each Statement	Statements about the HIV-PARSE Instrument
29	Depressing
25	Invasion of privacy
74	Repetitive
55	Asked about important parts of their lives
58	Might help others
61	Dealt with issues that should be studied as part of clinical trials for HIV treatment

In general, staff tended to overestimate how long it took patients to complete the form, how much assistance they needed, and the extent of patients' negative reactions to the form when compared to reports by patients. About 60 percent of the staff reported that some patients had difficulty reading the questionnaire and 30 percent reported that patients had physical difficulty writing answers to the questionnaire. About 60 percent of the staff members felt that the questionnaire was sad or depressing for patients, and 90 percent felt that it was repetitive. However, 80 percent of the staff felt that the questionnaire asked about things that should be studied as part of clinical trials.

[5]Response categories were "agree strongly," "agree somewhat," "neutral," "disagree somewhat," or "disagree strongly."

Reliability of the measures. In addition to concerns about the feasibility of administering these measures in a clinical practice setting, researchers are concerned about the reliability of the measures for patients with advanced HIV disease. Table 2.6 provides information about the reliability, central tendency, and variability of the scales in ACTG studies 081, 114, 116, 117, and 118, scored as described below. The reliability of the multi-item scales is very good, ranging from a low of .80 to a high of .90. In the Pfizer-CCTG MOPPS population, the reliability of the Pain and the Social Functioning scales was somewhat better because two-item versions of these scales were substituted for the one-item versions used earlier. However, in this smaller and sicker population, the reliability of other scales was somewhat lower. Specifically, the internal consistency reliability for the scales was as follows: Current Health Perceptions, .81; Physical Functioning, .86; Pain, .80; Energy/Fatigue, .80; Emotional Well-Being, .81; Cognitive Functioning/Distress, .89; Social Functioning, .76; and Role Functioning, .77. All these values are within the acceptable range for group comparisons.

SCORING RULES FOR MODIFIED HIV-PARSE QUESTIONNAIRE

Once the questionnaire is administered and the data are entered in machine-readable form and checked for errors, we recommend that the responses be scored as described below to aggregate them into usable form for analysis. Scoring is a two-step process. First, precoded numerical values are recoded per the scoring key given in Table 2.7. Note that all items are scored so that a high score defines a more favorable health state. In addition, each item is scored on a 0 to 100 range so that the lowest and highest possible scores are set at 0 and 100, respectively. Scores represent the percentage of total possible scores achieved. In Step 2, items in the same scale are averaged to create the ten scale scores. Table 2.8 lists the items averaged to create each scale. Items left blank (missing data) are not taken into account when calculating scale scores. Hence, scale scores represent the average for all items in the scale that the respondent answered and may include only one item as the entire score for the scale. Analysts may set a minimum number of items required before a scale score is calculated if they wish.

Table 2.6

HEALTH STATUS SCALE SCORES FROM THE HIV-PARSE EXPRESSED ON A 0 TO 100 SCALE WITH HIGHER NUMBERS BEING BETTER (HIGHER FUNCTIONING, LESS DISCOMFORT, ETC.)

Scale	Number of Items	Mean*	Standard Deviation*	Reliability
Current Health Perceptions	5	52	26	.88**
Physical Functioning	6	79	24	.87**
Pain	1	71	25	.64 - .81***
Energy /Fatigue	4	59	24	.89**
Emotional Well-Being	5	70	20	.86**
Cognitive Functioning/Distress	6	84	18	.90**
Social Functioning	1	77	27	0.74 - 0.84***
Role Functioning	2	77	29	.86**

*Based on 7352 responses in 1,856 participants.

**Cronbach's α calculated based on 1588 respondents. The maximum possible α is 1.0; scores of > 0.7 are considered acceptable for group comparisons (1, 2).

***Lower estimates based on the application of the Spearman-Brown Prophesy formula to MOS data; upper estimates based on a comparison of the test-retest stability at a four- to eight- (mean six-) week interval relative to that observed for the Current Health Perceptions scale.

PERCEIVED HEALTH INDEX

For many purposes it is preferable to summarize results of scale scores in a single index. Accordingly, we created a Perceived Health Index by regressing Current Health Perceptions on the six health status scores, yielding the following equation:

Perceived Health Index = 0.20*Physical Functioning + 0.15*Pain + 0.41*Energy/Fatigue + 0.10*Emotional Well-Being + 0.05*Social Functioning + 0.09*Role Functioning

where the numbers shown represent the proportional weight given each component scale scored over the range of 0 to 100 to yield a value for the index, which is also scored over the range from 0 to 100. The mean value of the Perceived Health Index in this population was 68 with a standard deviation of 19. The test-retest stability four to eight (mean = six) weeks for the Perceived Health Index was greater than that for the Current Health Perceptions scale (r = 0.73 versus 0.68). The internal consistency reliability of the index was 0.94 using conservative estimates for the reliability of the two single-item scales. A

more detailed discussion of the problems of computing and interpreting a single index score can be found in Bozzette, Duan, Berry, and Kanouse, 1994.

CONCLUSIONS

Overall, our results with the form in the Pfizer-CCTG trial are consistent with previous results, showing that it is quite feasible to incorporate measurement of health status and quality of life into HIV clinical trials of therapies, but that it is important to explain to both patients and clinic staff the purpose of the form, the reliability and validity of the measures (including why questions are seemingly repetitive), and how the resulting data will be used. The PARSE form is used in the context of a clinical trial that includes many tests and procedures unpleasant to receive or administer, but these provide valuable data for evaluating alternative therapies. Viewed in this context, the burden of completing the PARSE form is a reasonable trade-off against the value of providing reliable and valid measures of patients' experience with their health during the trial.

Table 2.7

STEP 1: RECODING ITEMS

Item Numbers	Change Original Response Category[a]	To Recoded Value of:
11,12B,12C,12D,12E,12I, 14,17	1 ---------->	100
	2 ---------->	75
	3 ---------->	50
	4 ---------->	25
	5 ---------->	0
12A,12F,12G,12H,22A-G	1 ---------->	0
	2 ---------->	25
	3 ---------->	50
	4 ---------->	75
	5 ---------->	100
13,20F,20J,20K	1 ---------->	100
	2 ---------->	80
	3 ---------->	60
	4 ---------->	40
	5 ---------->	20
	6 ---------->	0
15,16,18A,18B,18C,18D, 18E,18F	1 ---------->	0
	2 ---------->	50
	3 ---------->	100
20A,20B,20C,20D,20E, 20G,20H,20I,20L	1 ---------->	0
	2 ---------->	20
	3 ---------->	40
	4 ---------->	60
	5 ---------->	80
	6 ---------->	100

[a]Precoded response choices as printed in the questionnaire.

Table 2.8

STEP 2: AVERAGING ITEMS TO FORM SCALES

Scale	Numbers of Items	After Recoding Per Table 2.7, Average the Following Items:
Current health perceptions	5	11, 12A, 12B, 12E, 12H
Physical functioning	6	18A, 18B, 18C, 18D, 18E, 18F
Pain	2	13, 17
Energy/Fatigue	3	20G, 20I, 20K
Emotional well-being	5	20D, 20F, 20H, 20J, 20L
Cognitive functioning	3	20B, 20C, 20E
Social functioning	2	14, 20A
Role functioning	2	15, 16
Quality of life	7	22A-G
Will-to-function	5	12C, 12D, 12F, 12G, 12I

EXAMPLE: Items 14 and 20A are used to score the measure of social functioning. Item 14 has five response choices, and a high score (response choice 5) on Item 14 indicates extreme limitations on social functioning. Item 20A has six response choices, and a high score indicates the absence of limitations. Table 2.7 shows that responses 1 to 5 to Item 14 should be recoded to values 100, 75, 50, 25, and 0, respectively, and that for Item 20A, response choices 1 to 6 should be recoded to values 0, 20, 40, 60, 80, and 100, respectively. Table 2.8 shows that the two recoded items should be averaged to calculate the scale score for social functioning. If the respondent is missing one of the two items, the scale score will be equal to the recoded score for the non-missing item.

Scoring rules presented above are similar to those used in the RAND 36-Item Health Survey 1.0, (Hays, Sherbourne, and Mazel, 1993).

CHAPTER III

THE IMPACT OF PROPHYLAXIS AGAINST *MYCOBACTERIUM AVIUM* WITH RIFABUTIN, AZITHROMYCIN, OR BOTH ON HEALTH STATUS AND FUNCTIONING IN PERSONS WITH ADVANCED HIV INFECTION

INTRODUCTION

The mainstays of treatment for advanced HIV infection have been anitretroviral therapy and prophylaxis against certain opportunistic infections, particularly *Pneumocystis carinii.* In addition, regimens effective for primary prophylaxis against fungal, cytomegaloviral, and *Mycobacterium avium* disease have been developed. While the former two types of prophylaxis are not universally recommended, prophylaxis against *Mycobacterium avium* disease with rifabutin is. However, use of rifabutin prophylaxis is associated with a substantial risk of drug/drug interactions and treatment failure as well as substantial expense and bother of a daily dosing regimen. To address these concerns, Pfizer and the CCTG undertook a trial comparing standard daily rifabutin to a potentially more potent regimen, the combination of daily rifabutin and weekly azithromycin, and to a more convenient regimen, weekly azithromycin alone.

In this trial, rifabutin alone was the least effective regimen; combination therapy was both the most effective and the most toxic. Weekly azithromycin was superior in efficacy to rifabutin and was better tolerated than either of the other regimens. Of course, the relative safety and efficacy in preventing *Mycobacterium avium* infections are the prime considerations in the evaluation of these regimens. However, as these regimens are intended for chronic use, clinical and administrative decisionmakers will want to consider other factors in deciding which regimen to recommend. These will include the overall impact of the regimens on patients and on the utilization of health care.

Such information can be captured through the use of supplemental patient questionnaires, such as the HIV-PARSE instrument. Measures of health status such as those included in the PARSE have been shown to be sensitive to differences between clinically equivalent treatments in other settings (Bozzette, Kanouse, Berry, Duan, 1995). In HIV disease, several groups have shown that self-report survey instruments based on MOS general health and well-being measures can reliably capture health status (Wu et al, 1991; Hays and Shapiro, 1992). However, such assessments are confounded by attrition and mortality in HIV and other fatal diseases. In prophylaxis trials such as this one, attrition can be particularly important because, as in this one, such trials often suspend or sharply curtail data collection after the occurrence of a failure endpoint. Under these circumstances,

effective regimens not only miss being credited for avoided morbidity but are assigned relatively more HIV-related morbidity because patients remain on them longer. This is far from the ideal situation wherein prophylactic regimens are judged on the basis of the overall benefit and harm, including the total burden of morbidity and mortality caused by both the prophylactic treatment and the disease that it is intended to prevent.

This study examines information regarding health status, functioning, and utilization of health care reported by participants in the Pfizer-CCTG trial. It is an important limitation of this trial that, as in most such trials, follow-up was reduced after failing participants sought care for *Mycobacterium avium* disease. Longer term survey data are unavailable for such patients, and the unadjusted overall results reported will understate the impact of prophylaxis failures. Therefore, most of the results presented below should be interpreted to represent primarily the period preceding prophylaxis failure; some explicitly address only that period.

METHODS

Participants in the Pfizer-CCTG MOPPS study were persons with HIV infection and a CD4+ cell count of fewer than 100. Participants were randomized equally to receive rifabutin at 300 mg daily, azithromycin at 1200 mg weekly, or both. At entry and at each scheduled follow-up, participants received standard clinical and laboratory evaluations. In addition, participants were requested to complete survey questionnaires at baseline and at the 16-, 32-, 48-, 72-, and 96-week follow-up visits.

The HIV-PARSE survey instrument and the methods of analysis were essentially similar to those described previously. The survey includes questions on utilization, disability, work, and symptom impact as well as adapted versions of the MOS scales covering Current Health Perceptions (5 items), Physical Functioning (6 items), Pain (2 item), Energy/Fatigue (3 items), Emotional Well-Being (5 items), Cognitive Functioning/ Distress (3 items), Social Functioning (2 item), and Role Functioning (2 items). Also included were a 7-item Quality of Life scale and a overall Quality of Life categorical rating item: "Overall, how do you rate your quality-of-life?" The response scale for this item consisted of a horizontal line with the numbers 0 to 10 below and faces progressing from a frown to a smile above. On this scale, the 0 point was labeled "worst possible quality of life (as bad or worse than being dead)" while 10 was labeled "perfect health." This item is very similar to the 0 to 10 ladder used to obtain preferences and incorporate rating of death in the original studies with the Quality of Well Being scale. These scales and items have been shown to be highly reliable

and valid in this population (see Chapter II), to yield scores approximately normally distributed, and to correlate with health care utilization as well as with clinical and physiologic status. All scales were converted to a 0 to 100 scale with higher scores reflecting a more favorable health status in the standard fashion. When expressed this way, the scales scores can be understood to represent the percentage of the highest possible score. All scales were analyzed individually, and the MOS-based scales were also summarized into a perceived health index, which is a weighted average of individual scale scores.

Scale scores as well as data on utilization, disability, work, and income data were analyzed in separate subsets. In the first case, all data received were considered, and the data are censored only by death, attrition from the study, or completion of the study. While this is the most complete analysis, it is a biased assessment of the overall impact of prophylactic therapy because it understates the impact of prophylaxis failure. Accordingly, subset analyses of data received *before* a diagnosis of *Mycobacterium avium* disease were also performed. Specifically, the first subset analysis censors all survey data completed after cultures which were positive for *Mycobacterium avium* were obtained (ITT2); it is most useful for assessing the extent of the relative deleterious effects of prescribing the prophylactic regimens on persons without or prior to breakthrough *Mycobacterium avium*. The second subset also censors all data received more than 30 days after discontinuation of the study regimens (ITT1), and is most useful for examining the direct effects of the drugs in the regimens on the well being of patients who have not suffered breakthrough *Mycobacterium avium*.

Average values on all parameters were computed, and differences were assessed in a standard fashion. Simple and multiple ANOVA controlling for study week and baseline values were used in comparing continuous variables. Chi square tests and logistic regression controlling for study week and baseline values were used in comparing categorical variables. Additional analyses using non-parametric approaches for poorly distributed continuous variables were also performed.

One analysis to approach the question of overall impact in this limited dataset exploited the anchoring of the categorical rating to death in assigning scores of zero to persons who died. In the first case, the last score recorded for a subject was carried forward regardless of clinical status while, in the second, a score of zero was also assigned after a diagnosis of Mycobacterial disease.

To help handle attrition from mortality, drop-out, and reduction of data collection after the onset of breakthrough *Mycobacterium avium* infection, we adapted a novel approach to combining survival and health status data known as multistate survival analysis. In multistate survival analysis, estimates are made of the mean-time-above-threshold-state (MTATS) or mean total duration of time that persons have health status scores above a series of specified health status thresholds. When the threshold chosen is below the lowest recorded score, all ongoing (living) participants have scores above threshold and the MTATS is identical to the mean survival time. When higher thresholds are used, the two times are different because only the time above threshold rather than all time alive is considered in calculating the MTATS, and because all time above the given threshold is considered rather than just the time before the first drop to below the threshold as in standard survival analysis. The estimated MTATS can be calculated for many threshold scores spanning the range of observed values. The resulting family of MTATS values can be plotted on the Y axis against the range of thresholds on the X axis to yield a MTATS map. The MTATS map allows for visual inspection and comparisons of the time-quality experience of cohorts. In addition, the MTATS can be averaged across the range of thresholds to give the average mean time above threshold state (AMTATS), which can be interpreted as the typical time that a typical patient in the cohort spends above a typical threshold during the trial.

In this implementation, multiple imputation was used to correct for differential attrition. First, stacked ordered logistic regression was used on all data from all enrollees to develop an equation predicting transitions to a set of ordered clinical states--for example, a prediction of the proportion of persons transitioning fromalive without *Mycobacterium avium* infection to alive without *Mycobacterium avium* infection, alive with *Mycobacterium avium* infection, or dead. Next, ordered logistic regression was used on data from PARSE participants to develop an equation predicting health status thresholds from previous health index scores, study week, baseline characteristics, and clinical state. These equations were applied repeatedly to persons for whom clinical state and/or health status data were missing to create 250 datasets containing different imputed data. The differences in the datasets were induced to incorporate the error of prediction into the imputations in the following fashion. For each imputation in each dataset, a random number was drawn from the uniform distribution and applied to the cut points derived from the predictive mulitnomial logistic equations to impute the state to be assigned. Finally, to bound the effects of attrition after prophylaxis failure, analyses were repeated

assuming first that failure was equivalent to death and then that it had no effect on health status (i.e., treating *Mycobacterium avium* infection as a censoring event).

The significance of differences in the time-quality experience of patients as defined by the multistate survival analysis was assessed using both an area under the curve approach in which average TATS for the treatment groups was compared using ANOVA and by comparisons of individual break points on the curves.

All persons participating in the main study signed written informed consent forms approved by the relevant local Institutional Review Boards, and both this project and the HIV-PARSE survey instrument were reviewed by the RAND Human Subjects Protection Committee.

RESULTS

All study sites participated in this substudy. Of 723 study registrants, 677 of 723 (92 percent) completed baseline forms. Of the 645 persons who could be evaluated for effect of *Mycobacterium avium* infection, 502 (78 percent) completed a baseline and at least one other survey and are included in this analysis. Similarly, 1894 of 2065 (92 percent) surveys received, or 86 percent of the total number of expected number of forms were usable in this analysis. Survey participants were similar to nonparticipants and, among participants, persons randomized to the three arms were also demographically and clinically similar (Table 3.1a).

Persons randomized to combination therapy had more days of reduced activity than those assigned to monotherapy ($p = .04$), and there were lesser imbalances in other forms of disability days (Table 3.1b). That group also were more likely to have worked in a white collar occupation ($p = .005$), had a somewhat higher employment rate at entry and, among the employed, higher incomes and more days of missed work. The combination group also had more phone contacts with providers ($p = .03$), but tended to use somewhat less office and hospital care than those randomized to azithromycin. Baseline functional status and utilization were otherwise similar between the treatment groups.

Table 3.1a

BASELINE CHARACTERISTICS OF SURVEY BY NON-PARTICIPANTS AND PARTICIPANTS OF TREATMENT GROUP
Values given are percentages or means

	PARSE Participation		Among PARSE Participants		
	No	Yes	Azithromycin	Rifabutin	Both
N	130	502	173	162	167
Age	38	38	38	38	39
White	57%	63%	61%	61%	66%
African-American	20%	21%	23%	19%	19%
Hispanic	18%	15%	14%	17%	13%
Male	95%	95%	95%	96%	94%
CD4+	46	54	50	52	59

Overall Functional and Utilization Outcomes. Average health status scale scores declined little over the course of the trial (Figure 3.1). In addition, scale and index scores were remarkably similar across the treatment groups, with the average values for the health index summarizing most of these scale scores being identical in the three groups (Table 3.2). Statistically significant or near significant differences were observed in three of the scales, but no consistent pattern emerged, as the emotional well-being scale was highest in the combination arm (p for overall comparison = .01), the cognitive functioning/distress scale was highest in the rifabutin arm (p for overall comparison = .02), and the current health perceptions scale was highest in the azithromycin arm (p = .06). The largest of these differences was in the cognitive functioning scale, where the difference between the highest and lowest average scores was 4 points, or .16 standard deviations, far less than the .5 standard deviation which is generally characterized as a large difference. The overall count of self-reported symptoms as well as the number of symptoms reported to be at least moderately severe were slightly higher in the rifabutin arm (p = .08 and .03, respectively).

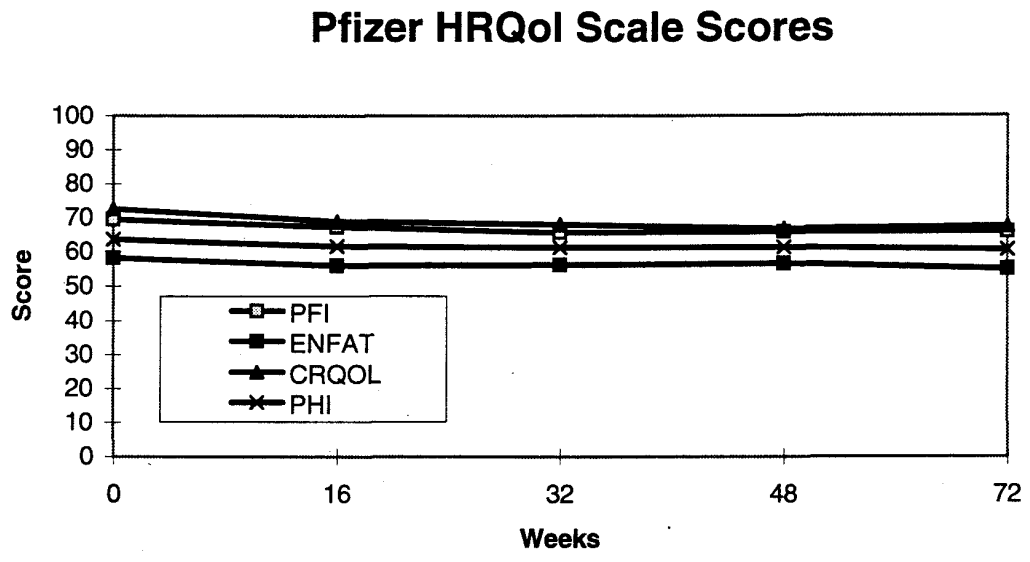
TABLE 3.1b

FUNCTIONAL CHARACTERISTICS OF PARTICIPANTS AT BASELINE
Values given are percentages or means

	Azithromycin	Rifabutin	Both
Proportion of candidates enrolled	34%	32%	33%
High school graduate	91%	90%	90%
College graduate	36%	40%	41%
White collar occupation (ever)***	71%	64%	80%
Physical Functioning	71	70	71
Role Functioning	68	70	69
Social Functioning	65	63	66
Mental Health/Emotional Well Being	68	68	72
Cognitive Functioning/Distress	80	77	80
Bodily Pain	68	63	66
Energy/Fatigue	59	58	58
Current Health Perceptions	49	46	45
Perceived Health Index	65	63	65
Quality of Life (scale)	51	50	51
Quality of life (categorical rating)	73	73	73
Hospital admission	4.1%	3.7%	3.0%
Average length of stay	6.9	5.7	4.5
Hospital days/month	0.3	0.2	0.1
Office visits/month	3.8	3.5	3.6
Home care visits/month	0.2	0.2	0.4
Provider telephone contacts/month**	1.3	1.6	2.6
Symptom count	18	18	18
Symptoms interfering moderately with functioning	4.2	4.8	4.8
Bed days (any)	45%	48%	44%
Bed days (number/month)	2.4	2.2	2.7
Reduced activity days (any)	55%	57%	57%
Reduced activity days (number/month)**	3.4	4.1	5.0
Days feeling less well (any)	57%	64%	67%
Days feeling less well (number/month)	3.8	5.2	5.6
Currently employed (full or part-time)	38%	41%	44%
Days of missed work/month (employed only)	2.4	3.1	4.8
Hours worked/week (employed only)	37	41	37
Hours worked/week (average of all enrolled)	13	15	16
Monthly earned income (employed only)	$2,973	$2,885	$4,132

All p values based on overall comparisons: * p = < .10, ** p = < .05, *** p = < .01.

FIGURE 3.1



Selected health-related quality-of-life scores over time for ongoing participants. PFI is physical functioning, ENFAT is energy/fatigue or vitality, CRQOL is categorical rating of overall quality of life, and PHI is the perceived health index, a summary measure for the MOS scales included in the PARSE. Note that changes in average scores over time are modest, probably because patients reaching endpoints are underrepresented and deaths are treated as missing scores.

Differences in disability generally favored the monotherapy arms. The proportion of persons with any bed days was lowest in the rifabutin arm ($p = .01$), while the both average number of bed days and days of reduced activity across all participants were lower in the monotherapy arms than in the combination arm ($p = .01$ and .02, respectively). Work variables such as average employment rate ($p = .03$) and average hours worked ($p = .02$) across all participants favored rifabutin despite baseline imbalances favoring the combination arm as patients in that arm experienced large drops in employment over the course of the trial.

Observed hospital use was similarly high across the treatment groups, with nearly 5 percent of the participants being hospitalized for an average of 8 days during an average month. Persons assigned to the combination arm made significantly more office visits ($p = .001$) and used more home care ($p = .04$). Borderline significant differences in telephone contacts disappeared after controlling for baseline imbalances.

Functional and Utilization Outcomes Prior to a Diagnosis of *Mycobacterium avium*. Values for the time prior to the onset of *Mycobacterium avium* infection under either definition of intent to treat were very similar to the overall data (Tables 3.3a, 3.3b). Differences generally ranged from 0 to 2 points in the last significant digit more favorable under ITT1 compared to all data and under ITT2 compared to ITT1. Tests of significance were also very similar in the three analyses, with the most notable differences being that formerly "significant" differences in emotional well-being and cognitive functioning/distress became borderline significant.

Inasmuch as only 1 or 2 percent of the data were obtained after prophylaxis failure, this extreme similarity is not surprising. It underscores the fact that this is primarily a study of the impact of treatment on persons who have not yet experienced breakthrough.

IMPACT OF BREAKTHROUGH *MYCOBACTERIUM AVIUM* INFECTION

Data were obtained from a small number of participants after a diagnosis of *Mycobacterium avium*, though none were obtained from those randomized to combination therapy. The paucity and selectiveness of the data make treatment group comparisons suspect, but the data are of interest nonetheless (Table 3.4). Energy and Pain scales are over half a standard deviation lower than in the pre-breakthrough data, with Role Functioning and Current Health Perceptions about a full standard deviation lower. Disability is much higher. Over 60 percent of these participants spent at least one day per month in bed, with a typical participant spending more than four days per month in bed and nearly a week per month of reduced activities because of health. Finally, employment was unusual.

Extended Analysis of Categorical Rating Data

Assigning scores of zero for death and carrying forward the last pre-breakthrough values for the categorical rating of overall quality of life gave overall mean scores of 64 for the azithromycin group, 64 for the rifabutin group, and 65 for the combination group ($p = .78$). Assigning scores of zero for both death and prophylaxis failure gave overall mean scores of 60 for the azithromycin group, 59 for the rifabutin group, and 64 for the combination group ($p = .0001$).

Table 3.2

HEALTH CARE, SYMPTOM IMPACT, DISABILITY, AND WORK
All figures are mean values per month including all data from patients who could be evaluated

	Azithromycin	Rifabutin	Both
Physical Functioning	66	66	67
Role Functioning	66	67	65
Social Functioning	61	63	61
Mental Health/Emotional Well Being***	68	70	71
Cognitive Functioning/Distress**	79	81	77
Bodily Pain	63	60	62
Energy/Fatigue	57	57	55
Current Health Perceptions*	47	46	44
Perceived Health Index	61	61	61
Quality of Life (scale)	48	49	50
Quality of life (categorical rating)	67	68	68
Hospital admission	4.9%	4.3%	4.8%
Average length of stay	8.9	7.7	7.4
Hospital days/month	0.38	0.32	0.34
Office visits/month***	3.0	2.8	3.6
Home care visits/month**	0.5	0.4	0.7
Provider telephone contacts/month	1.4	1.5	2.1
Symptom count	18	19	18
Symptoms interfering moderately with functioning	4.4	5.1	4.6
Bed days (any)***	45%	38%	47%
Bed days (number/month)***	2.1	2.1	2.7
Reduced activity days (any)	54%	49%	55%
Reduced activity days (number/month)**	3.4	3.7	4.2
Days feeling less well (any)	61%	62%	62%
Days feeling less well (number/month)	4.4	4.6	4.8
Currently employed (full or part-time)	33%	41%	34%
Days of missed work/month (employed only)	3.1	2.8	3.3
Days of missed work/month	0.8	0.6	0.8
Hours worked/week (employed only)	37	38	39
Hours worked/week (average of all enrolled)	11	14	12

All p values based on overall comparisons: * p = < .10, ** p = < .05, *** p = < .01.

Table 3.3a

HEALTH CARE, SYMPTOM IMPACT, DISABILITY, AND WORK PRIOR TO A DIAGNOSIS OF *MYCOBACTERIUM AVIUM* INFECTION
All figures are mean values per month including all data from patients who could be evaluated prior to breakthrough of prophylaxis (ITT2)

	Azithromycin	Rifabutin	Both
Physical Functioning	66	66	67
Role Functioning	66	67	65
Social Functioning	61	63	61
Mental Health/Emotional Well Being*	68	70	71
Cognitive Functioning/Distress*	79	80	77
Bodily Pain*	63	60	62
Energy/Fatigue	57	57	55
Current Health Perceptions*	48	47	44
Perceived Health Index	62	62	61
Quality of Life (scale)	49	50	50
Quality of life (categorical rating)	67	69	68
Hospital admission	4.6%	4.5%	4.8%
Average length of stay	9.3	7.7	7.5
Hospital days/month	0.39	0.33	0.34
Office visits/month**	3.0	2.8	3.6
Home care visits/month	0.5	0.3	0.7
Provider telephone contacts/month	1.3	1.5	2.1
Symptom count*	18	19	18
Symptoms interfering moderately with functioning*	4.3	5.0	4.6
Bed days (any)***	45%	37%	47%
Bed days (number/month)***	2.0	2.0	2.7
Reduced activity days (any)	53%	48%	55%
Reduced activity days (number/month)***	3.3	3.6	4.2
Days feeling less well (any)	60%	62%	62%
Days feeling less well (number/month)	4.2	4.2	4.8
Currently employed (full or part-time)**	34%	41%	34%
Days of missed work/month	0.8	0.6	0.8
Hours worked/week (employed only)	37	38	39
Hours worked/week (average of all enrolled)***	12	15	12

All p values based on overall comparisons: * p = < .10, ** p = < .05, *** p = < .01.

Table 3.3b

HEALTH CARE, SYMPTOM IMPACT, DISABILITY, AND WORK PRIOR TO A DIAGNOSIS OF *MYCOBACTERIUM AVIUM* INFECTION
All figures are mean values per month including all data from patients who could be evaluated prior to breakthrough of prophylaxis or up to 30 days after discontinuation of assigned drug (ITT1)

	Azithromycin	Rifabutin	Both
Physical Functioning	66	66	67
Role Functioning	66	67	65
Social Functioning	61	63	61
Mental Health/Emotional Well Being*	68	70	71
Cognitive Functioning/Distress*	79	81	77
Bodily Pain *	63	60	62
Energy/Fatigue	57	57	55
Current Health Perceptions*	48	46	44
Perceived Health Index	62	62	61
Quality of Life (scale)	49	50	50
Quality of life (categorical rating)	67	69	68
Hospital admission	4.8%	4.5%	4.8%
Average length of stay	9.3	7.7	7.5
Hospital days/month	0.38	0.32	0.34
Office visits/month***	3.0	2.8	3.6
Home care visits/month**	0.5	0.4	0.70
Provider telephone contacts/month	1.4	1.5	2.1
Symptom count*	18	19	18
Symptoms interfering moderately with functioning*	4.3	5.0	4.6
Bed days (any)***	45%	37%	47%
Bed days (number/month)***	2.1	2.1	2.7
Reduced activity days (any)	53%	48%	55%
Reduced activity days (number/month)***	3.4	3.7	4.2
Days feeling less well (any)	61%	62%	62%
Days feeling less well (number/month)	4.4	4.7	4.8
Currently employed (full or part-time)**	33%	41%	34%
Days of missed work/month	0.8	0.6	0.8
Hours worked/week (employed only)	37	38	39
Hours worked/week (average of all enrolled)***	12	14	12

All p values based on overall comparisons: * p = < .10, ** p = < .05, *** p = < .01.

Multistate Survival Analysis. The mean time above threshold state (MTATS) map incorporating multiple imputation to adjust for missing values after prophylaxis failure is our best estimate of the relative overall impact of prophylaxis as it attempts to compensate for attrition and to incorporate death. It suggests a slight (and statistically non-significant) advantage for azithromycin over combination and for combination over rifabutin over most of the range of observed quality-time (Figure 3.2). The area under the MTATS curves summarizes the overall quality-time experience and is expressed as the average of the MTATS or the AMTATS, which would be equal to 50 percent of the study duration in a stable population. By this measure, persons randomized to azithromycin had an average of 31.8 of 72 weeks (44 percent) with at least the typical health state while those randomized to combination had an average of 31.7 such weeks (44 percent) and those randomized to rifabutin had an average of 30.8 such weeks (43 percent). These differences do not approach statistical significance overall (average p for joint significance of intercept and slope of MTATS curves being different across 250 imputations = .32 for azithromycin versus rifabutin, .37 for combination versus rifabutin, and .80 for azithromycin versus combination), nor were differences at the individual thresholds significant (lowest p value for joint comparison at individual deciles = .42).

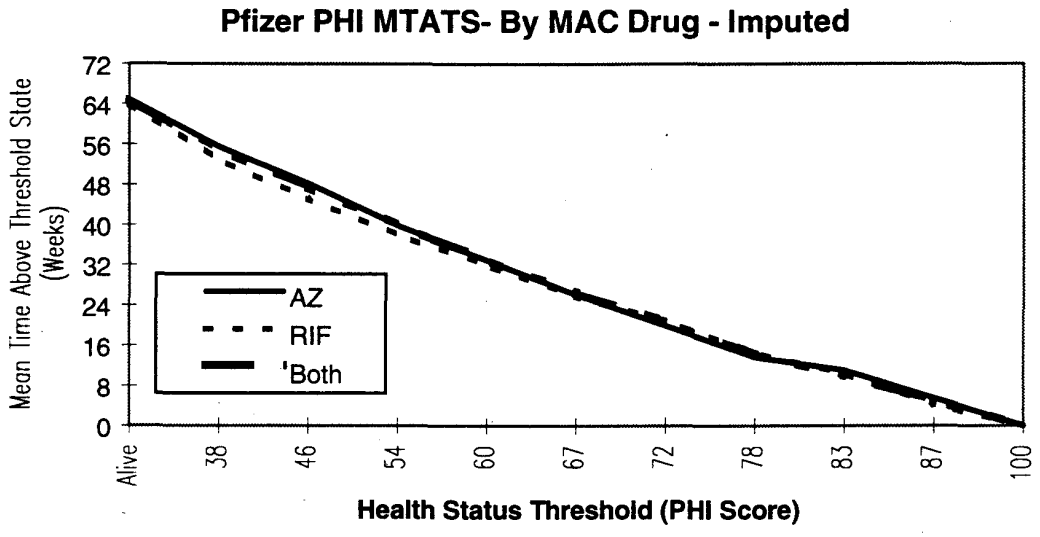
Further analyses were performed considering breakthrough a censoring event (i.e., as if a diagnosis of Mycobacterial disease had no consequences) or considering breakthrough a devastating event (i.e., as if a diagnosis of Mycobacterial disease was equivalent to death). The effect of these assumptions across all treatments arms is to raise estimates of the overall quality-time slightly in the first case and to lower estimates more substantially in the second (Figure 3.3). The effect of imputing no health consequences for breakthrough would be to increase the difference between azithromycin and combination but to decrease the difference between combination and rifabutin, as the AMTATS would be 31.9, 31.7, and 32.0 weeks for azithromycin, combination, and rifabutin, respectively (Figure 3.4; average p for joint significance of intercept and slope of MTATS curves being different across 250 imputations = .73 for azithromycin versus rifabutin, .75 for combination versus rifabutin, and .76 for azithromycin versus combination; lowest p value across individual deciles = .58). The effect of imputing devastating health consequences for breakthrough would be to decrease the quality-time by about two weeks for azithromycin, one-half week for combination therapy, and four weeks for rifabutin, giving estimated AMTATS of 29.8, 31.1, and 28.3, respectively (Figure 3.5; average p for joint significance of intercept and slope of MTATS curves being different across 250 imputations = .12 for azithromycin versus

Table 3.4

HEALTH CARE, SYMPTOM IMPACT, DISABILITY, AND WORK AFTER A DIAGNOSIS OF *MYCOBACTERIUM AVIUM* INFECTION
All figures are per month including all available data from patients who could be evaluated (ITT2)

	Azithromycin	Rifabutin	Both
N (data points)	16	19	0
Physical Functioning	61	57	
Role Functioning	45	56	
Social Functioning	51	61	
Mental Health/Emotional Well Being	65	69	
Cognitive Functioning/Distress	85	81	
Bodily Pain	51	61	
Energy/Fatigue	49	38	
Current Health Perceptions	23	29	
Perceived Health Index	53	50	
Quality of Life (scale)	39	35	
Quality of life (categorical rating)	55	53	
Hospital admission	12%	0%	
Average length of stay	1.5	-	
Hospital days/month	0.1	-	
Office visits/month	3.8	2.3	
Home care visits/month	0.7	0.8	
Provider telephone contacts/month	2.9	1.7	
Symptom count	18	19	
Symptoms interfering moderately with functioning	7	6.6	
Bed days (any)	69%	53%	
Bed days (number/month)	3.3	5.6	
Reduced activity days (any)	88%	58%	
Reduced activity days (number/month)	6.2	6.7	
Days feeling less well (any)	94%	63%	
Days feeling less well (number/month)	8.1	11.7	
Currently Employed (full or part-time)	6%	26%	

FIGURE 3.2



MTATS maps imputing consequences of breakthrough. The Y axis depicts the MTATS (mean time above threshold state) and the X axis depicts the range of observed health status scores/thresholds. The left-most points (at the origin of the X axis) indicate the time with at least the lowest observed health status, which is equivalent to the mean survival time. The right-most points indicate the time with health status higher than the highest observed score and are necessarily zero. Each tick mark indicates a decile of the scores observed at baseline; e.g., 46 is the 20th percentile, 67 is the median, and 83 is the 80th percentile. Solid line indicates the azithromycin group, dotted line indicates the rifabutin group, and the dashed line is the combination group.

rifabutin, .001 for combination versus rifabutin, and .16 for azithromycin versus combination; p for time above death or MAC = .02, lowest p across other individual deciles = .11).

DISCUSSION

The data from this substudy of the Pfizer-CCTG prophylaxis trial are best thought of as characterizing the health status, functioning, and utilization of the study population prior to the development of breakthrough disease. From that perspective, the results indicate some advantages for monotherapy over the combination of azithromycin and rifabutin. As subjective measures were generally quite similar, differences were seen primarily in disability measures, the use of office-based health care, and in greater increases in unemployment among persons randomized to combination therapy.

Because, by design, surveys were not obtained after a diagnosis of Mycobacterial disease was made, the study could not accurately and directly assess the relative overall impact of

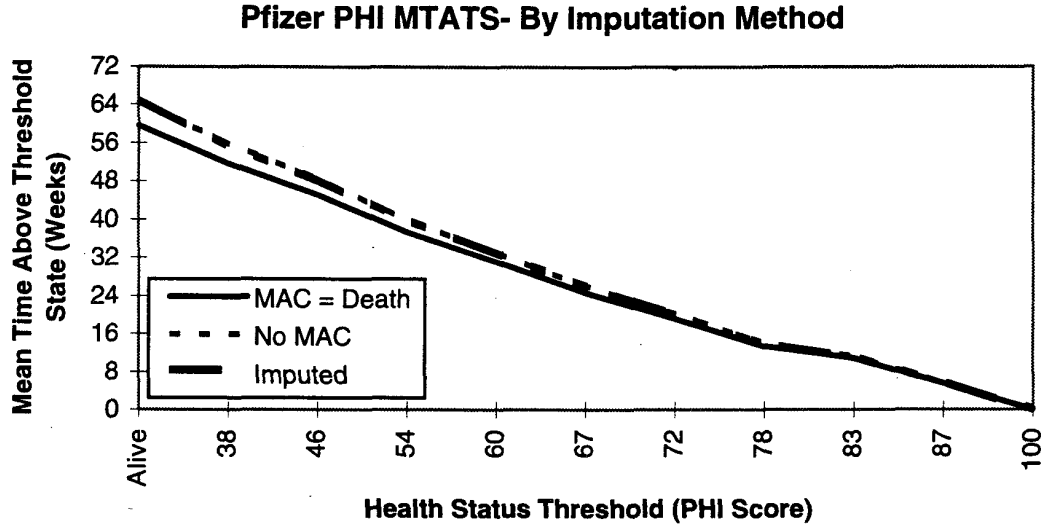
prescribing these regimens. However, the limited data obtained after prophylactic failure do provide insight into the devastating impact of early breakthrough disease. Compared with the pre-disease cohort, those with breakthrough disease had precipitous decreases in subjective health status, increases in moderately severe symptoms, a doubling of disability days, and none were employed. Effects on utilization were difficult to assess because of the small numbers. Overall, these results underscore the importance of determining impact of breakthrough disease. Clearly, assessing the impact of resistance in breakthrough isolates is critical: The overall benefit of prophylaxis is affected by the impact of breakthrough disease, which is itself determined in part by the sensitivity of breakthrough isolates and the success of treatment for resistant organisms.

Two analytic strategies coupled with a sensitivity analysis attempted to assess the question of overall impact in this limited dataset. The sensitivity analysis assumed that breakthrough was clinically trivial or clinically devastating. The first condition corresponds to a situation where breakthrough is easily detected and treated, whereas the second corresponds to a situation where is it is clinically devastating and resistant to treatment. These alternative assumptions were intentionally chosen to bracket the true effect of breakthrough disease. Differences between the overall health status or quality of life values under these two assumptions were not as extreme as the assumptions themselves, emphasizing that the differences in the total proportion of time spent with Mycobacterial disease were not very different in the three treatment groups.

The simpler of the strategies exploited the anchoring of the categorical rating to death in assigning scores of zero to persons who died. When we assumed that breakthrough had no consequences, the effect of the treatments was similar and reflected the similar outcomes in the pre-breakthrough cohort. When we assumed that breakthrough was equivalent to death, the most effective prophylactic regimen--combination therapy--resulted in the best overall quality of life.

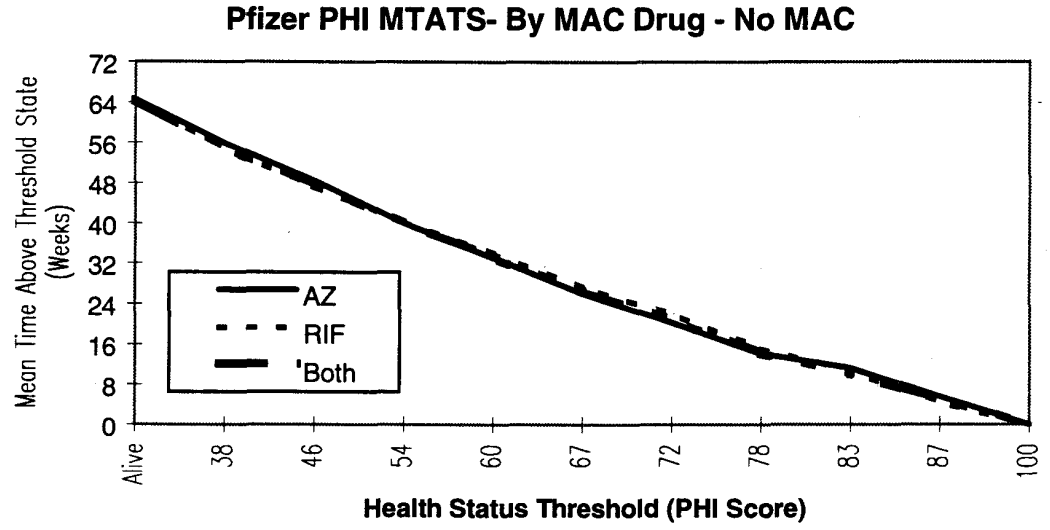
Our best estimate of overall benefit comes from the multistate survival analysis in which we employ a strategy of sequential multiple imputation and use survival analysis assumptions to handle attrition. In that analysis, the additional effectiveness of combination therapy did not translate into a superior overall benefit, and the point estimates of overall quality-time exceeded those for rifabutin alone only if one assumed breakthrough to be devastating--and then the differences were not statistically significant overall. Under the same circumstances, there was a trend toward superiority of

FIGURE 3.3



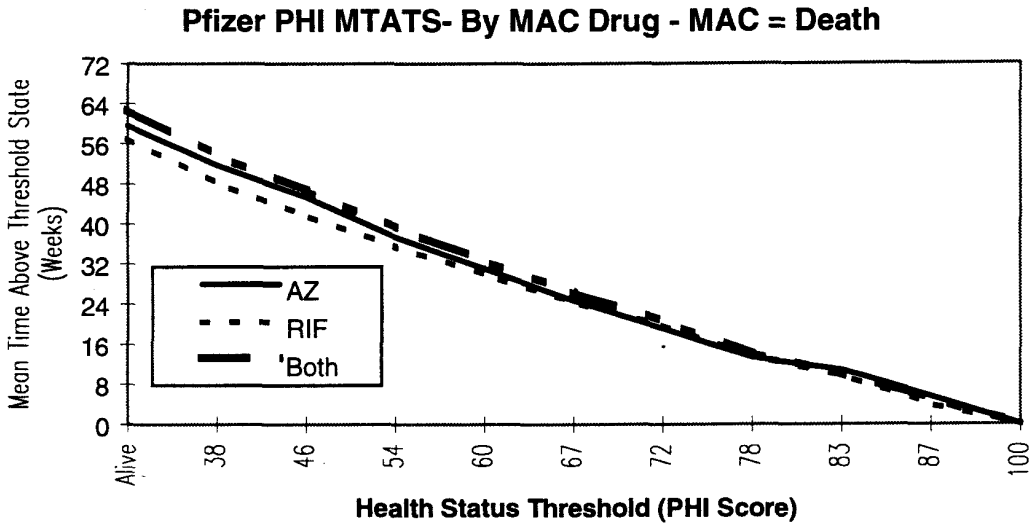
MTATS maps under different assumptions for consequences of breakthrough. Figure shows the effect of imputing the best estimate for the consequences of Mycobacterial disease (dashed line), assuming that the disease has no health consequences (dotted line), and assuming that it is equivalent to death (solid line).

FIGURE 3.4



MTATS maps under the assumption that breakthrough has no health consequences. Figure shows the effect of imputing under the assumption that Mycobacterial disease has no health consequences. Solid line indicates the azithromycin group, dotted line indicates the rifabutin group, and the dashed line, the combination group.

FIGURE 3.5



MTATS maps under the assumption that breakthrough is equivalent to death. Solid line indicates the azithromycin group, dotted line indicates the rifabutin group, and the dashed line, the combination group.

azithromycin over rifabutin. This makes a strong case for arguing that the excess cost, inconvenience, toxicity, and adverse effects on functioning are not outweighed by the superior efficacy of combination therapy.

In summary, certain features of the study design make an overall assessment of the relative value of these forms of prophylaxis difficult in this trial. These include the lack of close follow-up after breakthrough, a short overall duration of study treatment and follow-up, infrequent administration of survey instruments, and less than comprehensive measurement of symptoms specific to *Mycobacterium avium* disease. Nonetheless, the results obtained prior to a diagnosis of *Mycobacterium avium* disease demonstrate a personal cost to the use of combination prophylaxis; use of imputation strategies to compensate for differential attrition suggest that the superior efficacy of combination does not overcome that cost in leading to improved overall health-related quality of life, but similar imputation-based approaches should be undertaken for simple measures of functioning. These same analyses suggest that the pre-breakthrough impact of the two monotherapy arms was similar and that the superior efficacy of weekly azithromycin may translate into a greater overall benefit compared to rifabutin.

REFERENCES

Berry, SH, Bozzette, SA, Hays, RD, Stewart, AL, Kanouse, DE. Measuring Patient-Reported Health Status in Advance HIV Disease: HIV-PARSE Survey Instrument. MR-341-NIAID, RAND, 1994

Bozzette SA, Duan N, Berry S, Kanouse DE. Analytic difficulties in applying quality of life outcomes to clinical trials of therapy for HIV disease. Psychology and Health 1994; 9:143-156.

Bozzette SA, Hays RD, Berry S, Kanouse DE. A perceived health index for use in persons with advanced HIV disease: Derivation, reliability and validity. Medical Care 1994; 32:716-731.

Bozzette SA, Kanouse DE, Berry S, Duan N. Health status and function with zidovudine or zalcitabine as initial therapy for AIDS: A randomized placebo-controlled trial. JAMA 1995; 273:295-301.

Bozzette SA, Kanouse DE, Duan N, Berry S, Richman DD. The impact of zidovudine compared with didanosine on health status and functioning in persons with advanced HIV infection and a varying duration of prior zidovudine therapy. Antiviral Therapy 1996; 1:21-32.

Brown JB, Adams ME, "Patients as Reliable Reporters of Medical Care Process-Recall of Ambulatory Encounter Events," *Medical Care*, Vol. 30, No. 5, 1992, pp. 400-411.

Hays RD, Sherbourne C, Mazel RM "The RAND 36-Item Health Survey 1.0," *Health Economics*, 2:217-227, 1993.

Hays RD, Shapiro MF. An overview of generic health-related quality-of-life measures for HIV research. Quality of Life Research. 1992;1:91-7.

Kahn JO, Lagakos SW, Richman DD, Cross A, Pettinelli C, Liou SH, Brown M, Volberding PA, Crumpacker CS, Beall G, Sacks HS, Merigan TC, Beltangady M, Smaldone L, Dolin R, and NIAID and ACTG. A controlled trial comparing continued zidovudine with didanosine in human immunodeficiency virus infection. N Engl J Med 1992;327:581-7.

Stewart AL, Hays RD, Ware JE. Health perceptions, energy/fatigue, and health distress measures. In Stewart AL, Ware JE, eds. Measuring functioning and well-being: The Medical Outcomes Study approach. Durham and London: Duke University Press, 1992: 143-172.

Wu AW, Rubin HR, Mathews WC, Ware JW, Brysk LT, Hardy WD, Bozzette SA, Spector SA, Richman DD. A health status questionnaire using 30 items from the medical outcomes study. Preliminary validation in persons with early HIV infection. Medical Care 1991;29:786-798