



Published in final edited form as:

*AIDS Behav.* 2013 January ; 17(1): 127–141. doi:10.1007/s10461-012-0272-4.

## **MACH14: A Multi-Site Collaboration on ART Adherence Among 14 Institutions**

**Honghu Liu,**

Division of Public Health and Community Dentistry, School of Dentistry, University of California, Los Angeles, CA, USA

**Ira B. Wilson,**

Public Health Program, Warren Alpert School of Medicine, Brown University, Providence, RI, USA

**Kathy Goggin,**

Department of Psychology, University of Missouri – Kansas City, Kansas, MO, USA

**Nancy Reynolds,**

School of Nursing, Yale University, New Haven, CT, USA

**Jane M. Simoni,**

Department of Psychology, University of Washington, Seattle, WA, USA

**Carol E. Golin,**

School of Medicine, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

**Marc I. Rosen,**

School of Medicine, Yale University, New Haven, CT, USA

**Robert Gross,**

Epidemiology Center for Clinical Epidemiology and Biostatistics, School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

**Glenn Wagner,**

RAND Corporation, Bethel, CT, USA

**Robert H. Remien,**

HIV Center for Clinical and Behavioral Studies, New York State Psychiatric Institute and Columbia University, New York, NY, USA

**Neil Schneiderman,**

Department of Psychology, University of Miami, Coral Gables, FL, USA

**Judith A. Erlen,**

Department of Health and Community Systems, School of Nursing, University of Pittsburgh, Pittsburgh, PA, USA

**Julia H. Arnsten, and**

Departments of Medicine, Psychiatry, and Epidemiology, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY, USA

**David R. Bangsberg**

---

© Springer Science+Business Media, LLC 2012

Correspondence to: Honghu Liu, [hhliu@ucla.edu](mailto:hhliu@ucla.edu).

The study was conducted for the MACH14 Group.

Ragon Institute of MGH, MIT and Harvard, Massachusetts General Hospital Center for Global Health, Harvard Medical School, Boston, MA, USA

Honghu Liu: hhliu@ucla.edu

## Abstract

The integration of original data from multiple antiretroviral (ARV) adherence studies offers a promising, but little used method to generate evidence to advance the field. This paper provides an overview of the design and implementation of MACH14, a collaborative, multi-site study in which a large data system has been created for integrated analyses by pooling original data from 16 longitudinal ARV adherence studies. Studies selected met specific criteria including similar research design and data domains such as adherence measured with medication event monitoring system, psychosocial factors related to adherence behavior, and virologic and clinical outcomes. The data system created contains individual data (collected between 1997 and 2009) from 2,860 HIV patients. Collaboration helped resolve the challenges inherent in pooling data across multiple studies, yet produced a data system with strong statistical power and potentially greater capacity to address key scientific questions than possible with single-sample studies or even meta-analytic designs.

## Keywords

HIV/AIDS; Adherence; MEMS Multi-site Antiretroviral medication; Individual participant data

## Introduction

Adherence to antiretroviral therapy (ART) has proved to be vital to its success, yet research on ART adherence has been stymied by studies of self-reported adherence with small samples. Highly active antiretroviral therapy (HAART) has led to a striking decrease in both morbidity and mortality related to HIV/AIDS infection [1–4]. Research indicates that adherence to HAART is the strongest predictor of viral suppression, resistance, disease progression and death [4–6]. To provide optimal HIV care, clinicians and researchers need to understand the predictors of adherence to HAART and its influence on HIV health outcomes. Different measures and aspects of HAART adherence have been examined in a host of studies for more than a decade. This research has substantially improved our understanding of HAART adherence, however our knowledge remains incomplete. Different approaches to adherence measurement have made it difficult to compare findings across studies and limited researchers' ability to explain the inconsistencies observed.

Self-reported adherence measures have been most commonly used. They are valued for their convenience and practicality. However, scale items and methods of summarizing and analyzing data have varied considerably from one study to another, and self-reported measures are criticized for being imprecise in comparison to other types of measures because self-report is subjective and its accuracy depends on the respondent and factors such as the accuracy of memory. Although self-reported adherence usually correlates with virologic outcomes in the expected direction, it is generally found to overestimate adherence [4, 7–13].

The medication event monitoring system (MEMS), which electronically records the date and time of pill bottle openings, has the potential to provide a more objective measure of adherence when used correctly. While MEMS has been associated with virological response, studies employing MEMS are usually limited in sample size [14–16]. Consequently, most studies measuring adherence with MEMS have not had sufficient statistical power to conduct in-depth multivariate analyses of factors in relation with adherence, how these

relationships change over time, or outcomes of adherence beyond that of viral suppression such as patterns of resistance, changes in CD4 cell counts, disease progression or death.

To address these limitations, the Multi-site Adherence Collaboration in HIV among 14 institutes (MACH14) study (<http://www.mach14.med.ucla.edu/index.htm>) was initiated. This ongoing, NIMH-funded study (R01MH 078773) pools MEMS and other clinically relevant data from 16 different longitudinal HAART adherence studies to create a large, diverse HIV data system that can answer questions that are difficult or not feasible to address using any small, single study. Compared with individual subject-level studies, pooled data systems are more informative, have stronger statistical power, and can assess sources of variability across studies as well as between individual subjects. However, because the MACH14 studies were designed and executed independently, there are a number of challenges associated with conducting individual subject-level data meta-analysis. These challenges include non-uniform study entry criteria and varying data timing (e.g., some studies collected viral load (VL) at specified intervals and others used VLs collected in the course of routine clinical care), the use of different instruments and different frequency of measurements for key constructs and missing data—including structurally missing data (e.g., data never collected, collected once, collected irregularly) and other missing data (e.g., subject skipped a visit/item on a questionnaire, or dropped out.) Proper analytical and statistical methods need to be employed to adequately address these issues so that valid and effective results can be obtained from this unique and powerful pooled HIV data system.

In this paper, we present the process and methods for developing this first of its kind collaboration in medication adherence research to help inform future studies that might benefit from incorporating such methods. We also present the characteristics of individual studies involved and of their participants and report descriptive statistics of the currently available domains of the pooled data system. Finally, we discuss additional challenges and future research directions.

## Methods

### Development of MACH14 Collaboration and Study Selection

We assembled a list of potential collaborators from the NIH search engine RePORTER (<http://projectreporter.nih.gov/reporter.cfm>) to identify studies that used MEMS to study antiretroviral (ARV) adherence in the United States. Eligible studies were required to have: (1) a longitudinal study design with at least 3 repeated measurements; (2) MEMS adherence data; (3) VL and clinical outcomes; and (4) psychosocial and behavioral measures. To expand our search, we also submitted a formal invitation letter via an HIV adherence email list to a large pool of HIV adherence researchers. In the end, collaborators from 14 different institutions were identified. Each had data to contribute from at least one study that met the eligibility criteria; two of the collaborators had data from two studies, for a total of 16 studies. MACH14 was funded with a 5-year R01 from NIH/NIMH and commenced in June, 2007.

### Structure and Management of MACH14

The investigators at each of the 14 institutions worked closely with the Statistical and Data Coordination Center (S&DCC), which is comprised of methodologists, biostatisticians, analysts and programmers at the University of California Los Angeles (UCLA). The S&DCC was responsible for overseeing data transfer, merging and cleaning individual datasets, and collaborating with sites and conducting statistical analyses. Eight subcommittees were established to provide expertise in several domains such as MEMS and

Self-Reported Adherence, Drug Resistance, VL and ARV Medications, Substance Abuse and Adherence, Adherence Interventions, and Psychosocial Impact. A publication committee that includes all individual site PIs was also established. The full MACH14 group conducts monthly conference calls and an annual in-person group meeting; subcommittees have weekly, bi-weekly or as-needed meetings to discuss analyses and oversee paper-writing.

## Data Preparation and Transfer

**Overview of Data Management Procedures**—We identified the common domains of data from the 16 studies: subject demographic characteristics, MEMS adherence, VL and CD4 levels, drug resistance, self-reported adherence, as well as psychosocial, clinical, and other domains such as ARV regimen, substance abuse, sexual and behavioral risk, and adherence intervention. Data systems for each of these domains were first prepared at individual sites and then transferred to the S&DCC electronically. We chose to maintain large data systems separately for each domain to provide flexibility in data management and manipulation by allowing them to be merged with the other relevant datasets to form different analytical data files that meet the needs of various research questions and associated analytic plans. The data gathering and management process comprised the following four steps.

**Step One: Creation of Data Preparation Protocols**—We first created an Overall Data Preparation Protocol (ODPP), which stipulates guidelines and principles for preparing and transferring data, including procedures to maintain data security. The ODPP specified the domains of data systems to be transferred and the steps needed to be taken at both the site and S&DCC levels.

**Step Two: Data Preparation at Each Study Site**—We also created a data information document (DID) for each domain of the data to be transferred. The DID specified the data fields required for each subject, including variable names and attributes (e.g., numeric or character), coding, and units used. A data preparation protocol (DPP) accompanied each DID and described the procedures to be followed by every site, including the uniform screening and cleaning process algorithm for MEMS and other data, and the common definitions of created measures (e.g., race categories). Codebooks containing standard variable names and coding specifications were produced for each domain. Data preparation at each site included the conduct of logic, range/outlier, and error checks followed by the correction of errors discovered. Wherever possible, raw data, not derived variables, were requested from sites, so that standardized derived variables could be created. For example, we requested raw data on times of pill cap openings for each individual, not aggregated or summary adherence data. Each site transferred its prepared datasets to the S&DCC.

Every site and study complied with the data security procedures principles. An analytical ID for each subject was generated using site-specific scramble functions (accessible only to PIs). This encrypted ID is the only subject identifier in the analytic files.

**Step Three: Data Merging**—The S&DCC merged the data received from the sites for each domain and then re-checked the merged data for errors, collaborating with each site to correct errors, if any.

**Step Four: Analytic Variable Creation**—Once the raw data were received, merged and cleaned, the S&DCC created derived analytic variables. We created for each subject a summary “percent adherence” variable representing the number of valid pill bottle openings divided by the number of total openings that would have been expected (based on regimen

information entered into the MEMS program). If a person's calculated adherence was over 100 %, it was truncated to 100 % in analysis.

The final merged data system contains pooled data from 16 individual studies at 14 institutions across 12 states with more than 50 different domains of measurements and more than 260 common measures of subject demographics, MEMS, VL and CD4, drug resistance, self-reported adherence, psychosocial/behavioral factors, substance abuse, sexual risk, and adherence intervention.

## Data Analysis and Statistical Approaches

**Data Collection Intervals**—Some studies involved regular study visits for data collection. Among these studies, the interval between study visits varied from every month to every 3 months. Other studies tied study visits to receipt of clinical care or were “unannounced” (i.e., for pill counts). Bangsberg and colleagues [4, 17] developed the unannounced home-base pill count protocol, which relied on subjects to count their own pills and report the values to an assessor. Some studies collected VL data at specified intervals, and others used VLs conducted in the course of subjects' routine clinical care. Thus, for each analysis, decisions were made about how to account for these variations to allow for the maximum number of studies and subjects to be included in each analysis.

### **Different Measures Used and Amounts of Data Collected for Key Measures**

Since each of the 16 studies was designed and executed independently of the others, multiple measures were used across studies to assess the same construct. For example, the beck depression inventory (BDI), the Patient Health Questionnaire (PHQ-9), and the Center for Epidemiologic Studies Depression Scale (CES-D) were used to measure depression in different studies. We created common depression variables by standardizing the different measures using parameter estimates from normative population data for each scale and then converting them to z scores or categorizing the different measures into common levels of depression severity. Similarly, for the sub-domain of anxiety across the studies, the Zung Anxiety Scale, the Beck Anxiety Index, the State-Trait anxiety Inventory (STAI), and the Brief Symptom Inventory (BSI) were implemented in different studies.

Different amount of data were collected for some key measures across the studies. To resolve the heterogeneity and derive common measures that were suitable for analyses, we created broad classifications that could incorporate most measures. For example, regarding substance abuse, we developed measures, such as “ever abused drugs or alcohol”, to fit the situations across the studies.

**Viral Load, CD4 and Drug Resistant Data**—The key measures in the VL data consist of subject ID, date of blood drawn, type of lab (e.g., university clinical lab or commercial lab) and actual lab name, data source(e.g. phlebotomy at study visit, or medical record abstraction), actual VL reading with lower and upper limit, and lab assay methods such as quantitative polymerase chain reaction (PCR) or branched-chain DNA (BDNA).

Data on CD4 counts include date of blood drawn, type of lab (i.e., university clinical lab or commercial lab), and actual lab name and CD4 counts. The drug resistance data contain measures about gene (protease and reverse transcriptase), codon, substitution, and mutation.

**Clinical Data**—Clinical measures include lowest ever absolute CD4 count in units of cells/mm<sup>3</sup>, highest VL in record in copies/ml or log<sub>10</sub> copies/ml, CDC stage A, B, C at time of entry into study, duration of known HIV status at time of enrollment, ever had an HIV-related infection or malignancy or complication (e.g., wasting), currently use or ever used any kind of alternative or complementary medications, treatments, or supplements such as

vitamins, Chinese herbs, dinitrochlorobenzene (DNCEB), Qigong, ever taken or currently on PCP prophylaxis, and ever taken or currently on TB prophylaxis.

**Psychosocial Data**—Psychosocial measures include the following sub-domains: self-efficacy, anxiety, depressive symptoms, social support (general and medication-specific support) and perceived stress, coping, HIV and ARV related physical symptomatology, reasons for non-adherence, beliefs about medications, health surveys of health functioning and quality of life, adherence motivation as well as subjects' perceptions of physicians' competence and concern.

**Adherence Calculation Using MEMS Data**—The key measures in MEMS data system include subject ID, MEMS monitor ID, drug name, prescribed doses per day, the opening event day and time, time since last opening, and the start and stop date of the monitoring period. Five studies also collected information on the non-monitored period of MEMS to indicate when the electronic monitoring was interrupted due to special events and issues (e.g., for a hospitalization; for imprisonment).

MEMS data were used to calculate medication-specific and overall (across simultaneously prescribed medications) percent adherence. Overall adherence was calculated as the average adherence across the multiple ARVs taken. Percent adherence is calculated as the number of doses recorded by MEMS over the number of prescribed doses for a specific time period. We also calculated other adherence measures such as dose timing error.

**Statistical Approaches**—Given that all MACH14 studies had longitudinal designs, repeated measures mixed effects models (RMMEM) [18] were used as the backbone approach to model outcomes (e.g., changes in adherence, VL and CD4 over time). RMMEM has the advantages not only of being able to model global fixed effects (e.g., effects of gender, race or education on adherence), but also of being able to model the individual random effects (e.g., changes of adherence over time at subject level). Other modeling approaches, such as growth curve analysis, generalized estimating equation (GEE), and survival analysis methods were also used to analyze the data as needed for particular research questions. Non-linear relationships between outcome measures and predictors and covariates were considered and evaluated using non-linear models or spline techniques such as cubic splines.

## Results

### Challenges Encountered

The S&DCC merged the data received from the sites and re-checked the merged data for errors. Despite detailed specifications in the DIDs, error checking of the merged data for each of the domains usually revealed many different challenges. Because each of the 16 studies in MACH14 had different objectives, data collection schedules and analytic plans, a significant amount of effort was spent to identify and resolve a variety of data challenges.

Among all the MACH14 data, the MEMS data system was the most complex and required the most attention. The MEMS DID specified three files associated with MEMS data: file 1—MEMS data (including subject ID, MEMS monitor ID, drug name, number of prescribed daily doses, bottle opening date and time, time since last opening); file 2—start and stop dates for which MEMS data were monitored; file 3—MEMS modifications, that is, periods for which MEMS data are “missing” or non-monitored.

MEMS data system was cleaned by addressing non-use or malfunction of the MEMS caps; wrong MEMS ID associated with the start and stop dates; time zone errors introduced by

MEMS devices; missing data introduced by file corruption error; tracking errors, and erroneous large doses per day.

### Overall Descriptive Information

Table 1 summarizes the general information of institution, study name, interventional/observational study, intervention description, type and length for interventional studies, project duration, number of subjects, and length of follow-up of each study. These 16 NIH-funded HIV adherence studies from 14 institutions and 12 states were conducted between 1997 and 2009, and included a total of 2,860 HIV subjects. The mean length of subject follow-up was 16.5 months, ranging from 3 to 60 months. The sample size for each study ranged from 76 to 404.

The number of measures for key MACH14 variables by study are shown in Table 2. Of the 2,860 subjects, 2,498 (87 %) have available MEMS data, with a total of 478,242 individual MEMS openings documented; 2,776 (97 %) subjects have VL data with a total of 16,250 individual measurements. CD4 count data are available from 15 studies. There are 12,532 CD4 measurements in 2,575 subjects. Five studies collected drug resistance data using genotype or phenotype testing at virologic failure. Of the total 2,860 participants, 236 have tested for drug resistance with a total of 1,531 drug resistant data points collected.

### Demographic Data

Subject characteristics for the pooled data system are shown in Table 3. Mean (SD) age is 41(8) years, with a range of 18–72 years. Thirty-two percent are female. Nearly half (48 %) are Black/African American, followed by 29 % Caucasian/White, 15 % Latino/Hispanic, and 9 % “Other”. The majority (63 %) has high school diploma, 69 % were unemployed, and 58 % earned less than \$10,000 yearly. Regarding sexual orientation, 41 % are heterosexual, 34 % homosexual, 14 % are bisexual, and the rest (11 %) are not sure. In total, 476 (18 %) were ARV naïve subjects, who were initiating an ARV regimen.

### MEMS Data

Table 4 shows the distribution of ARVs monitored, including over 30 different ARV medications from all ARV drug classes. The most frequently monitored medications are Nelfinavir (NFV), Lopinavir/Ritonavir (LPV), Efavirenz (EFV), Combivir (CBV), and Indinavir (IDV), and average person months of follow-up for these medications are 4.62, 4.47, 5.89, 3.82, and 4.15, respectively. Figure 1 shows the box plots of adherence by month for the first 12 months. In general, median adherence tended to go down from month 1 to month 10, and then rose at months 11 and 12. The rise in adherence at months 11–12 could be due to improved adherence in some individuals or less data at those time points from less adherent individuals.

### Viral Load and CD4 Data

Table 5 shows distributions of the recorded VL and CD4. The first recorded VL for each subject ranges from undetectable (<50 copies/ml) up to 6.8 log<sub>10</sub> copies/ml, with 78 % detectable (≥ 50 copies/ml), and the mean of the first VL is 3.32 log<sub>10</sub> copies/ml with vast variation. After averaging across all the VLs that each subject had, the mean VL is 2.87 log<sub>10</sub> copies/ml, indicating that VL level becomes lower after the first VL. For the first recorded CD4 count, the mean is 359. After averaging across all the CD4 that each subject had, the mean CD4 is 396, indicating that CD4 level becomes higher after the first CD4.

## Substance Abuse Data

A total of 9,996 individual substance and alcohol abuse data points were collected for 2,820 subjects over time across the 16 studies. These data, though not necessarily in all the 16 studies, included substance use, alcohol use, substance abuse treatment such as Methadone treatment, overnight stay, symptoms withdraw from drug/alcohol, and legal problems (e.g., recent charges, arrests, probation and parole). Substances assessed included cocaine, sedatives (hypnotics), stimulants, psychedelics/hallucinogens, marijuana, and heroin. About one quarter (498/2,018) of subjects had current or recent substance abuse treatment, such as Methadone treatment, or residential treatment; 6.2 % (25/402) injection drug users; and 8.1 % (44/546) subjects have legal problems in the past year. Of all the available 2,610 substance/alcohol users, there are 47.7 % (1,172/ 2,459) alcohol drinker and 18.7 % (219/1,172) binge drinker in the past month. Of all the available substance users, 24.5 % (498/2,031) subjects use marijuana, 18.1 % (368/2,034) subjects use heroin, 13.4 % (294/2,198) use cocaine, 8.3 % (92/1,112) subjects use sedatives, 6.4 % (118/1,857) subjects use stimulants, 1.8 % (23/1,266) subjects use psychedelics or hallucinogens, and 1.7 % (25/ 1,501) subjects use ecstasy.

## Clinical and Psychosocial Data

Clinical data distributions are shown in Table 6. Psychosocial measures include self-efficacy, anxiety, depressive symptoms, general support, medication specific support, perceived stress, coping symptom, reasons, beliefs, health survey, motivation, and competence and concern. To resolve the issue of different measures for the same sub-domain of anxiety, a common anxiety measure across the studies was created by re-scaling the different anxiety measures to a z score using the mean and standard deviation of anxiety obtained from a population norm. To create the comparable measures, we mapped the different measures of coping onto the Brief Coping scale on a 0–100 scale.

## Intervention Programs

Interventional/observational study information is shown in Table 1. Of the 16 studies, 12 were intervention studies and 4 were observational studies. For the intervention studies, the majority had a true control arm that was mainly standard care programs. Two of the 12 interventional studies had intervention arms only (one had two intervention programs (supportive counseling and supportive counseling with a review of MEMS data and prizes for MEMS-verified medication-taking)) and the other had a factorial design with two intervention arms (“pager”, “buddy” for dose reminders to combat “forgetting” and relieve participant burden) evaluated in a factorial design. One study had a cross-over design. Although there were different intervention programs among the studies, all programs were designed to promote adherence to ARV medications.

## Discussion

In clinical trials, meta-regression of aggregated data is the usual approach for relating sources of variation in treatment effects to specific study characteristics. However, study-level analyses can lead to biased assessments, and use of aggregated summary values has many limitations for explaining the heterogeneity [7, 19–21]. Meta-analyses with individual participant data (IPD) gathered from the constituent studies allows researchers to perform subgroup analyses not conducted by the original studies and to add new information to the data sets by using different statistical methods [22].

A major challenge of this super-meta approach of pooling individual data of multiple trials together is to balance the internal differences between studies in order to build a virtual large trial. The MACH14 approach differs from the usual individual subject level meta analysis in



that most of the individual level meta analyses use drug trial data, while what we are attempting is much more complicated, and requires many more post hoc alignment and adjustment. The pooled MACH14 data system contains original measures at individual subject level, and has the capacity and strength to evaluate heterogeneity across studies, the relationships between subject characteristics and adherence as well as the association between adherence and virologic and clinical outcomes.

The pooling of individual trials in MACH14 has created a unique and powerful HIV data system with 2,860 subjects to address questions that are difficult for the individual studies. The data system represents the largest HIV anti-retroviral MEMS adherence data system ever assembled, which allows for specific analyses in sub-populations such as women, specific types of substance users, and particular regimens. The data system allows for better adjustment for potential confounding and the examination of interactions that are not permitted by smaller studies. A further advantage of pooling trials together is that we can use unified statistical methods which facilitate comparisons across all sites. Moreover, if consistent patterns across sites are identified, results could be generalized by pooling data from different geographical regions, which increases external validity.

A number of challenges associated with performing an individual level meta analysis arise since MACH14 studies use a wide variety of methods. Although MACH14 required that all included studies collected MEMS adherence and virologic data, and all had longitudinal designs with at least 3 repeated measurements, original study entry criteria were not uniformed. For example, not all studies have drug resistance data. Different studies could collect VLs at different intervals, some with specified intervals, and others did so in the course of subjects' routine clinical care. Furthermore, different measures were used for some key information such as depression and different amounts of data collected for constructs such as substance use across the different studies.

The process of transferring, cleaning and merging data has required close cooperation among investigators from multiple sites. Although strict selection criteria were used to ensure that the 16 selected studies were generally compatible, each individual study used different measures of common concepts such as depression and substance abuse. This heterogeneity led to a number of complications. Most notable is structural missing data due to variation in variables assessed across sites (e.g., anxiety data in only 1,388 participants), which necessarily requires some entire studies to be excluded from analyses. The mix of observational studies with intervention studies is another potential weakness of this pooled data system. For some types of analyses, data from observational studies and usual care condition arms can be combined to form a non-intervention group. For some other types of analyses, data from interventional studies can be grouped together to study intervention effect.

The reasons and nature of missingness in this pooled dataset are widely diverse. Data could be missing because a study collected data irregularly, a subject missed a visit, skipped an item on a questionnaire, or even dropped out of the study. In some cases, data are missing due purely to design and structural reasons. For example, not all studies assessed depression. For those that did, due to differential data collection schedules, some studies collected depression with different intervals. These structural differences in the dataset create problems because each study had different eligibility criteria that are related to adherence. In addition, these structural differences further complicate statistical analyses and modeling, especially repeated measures analyses and modeling because of varying intervals and availability of data points. Data were also missing for nonstructural reasons, such as missed study visits by patients, skipped items on a questionnaire, or were lost to follow up. Naïve methods for handling missing data (e.g., complete case analysis, pairwise deletion, and mean

substitution) can introduce substantial bias, reduce precision of estimates, reduce study power, and thus may lead to invalid study conclusions.

For MEMS missing data, we can create two versions of the hierarchical percent adherence measure: one calculated by MEMS data only and one calculated by MEMS and enhanced by self-reported adherence data. The latter version incorporates self-reported adherence into the calculation to reduce the amount of missing values [7]. As self-reported adherence likely overestimates a subject's true adherence level, repeated measures calibration models are fitted to calibrate self-reported adherence to the metric of MEMS [7, 23, 24].

Handling missing data is an important but very complex issue. For structurally missing data, imputation strategies may not be appropriate (e.g., imputation is not a practical approach for obtaining missing anxiety data if a study did not assess anxiety or did not assess it at certain time points). For other missing data situations, imputation or the use of missing indicators in analysis may be useful. However, the frequency and degree, as well as the mechanism of missingness (the extent to which missingness is at random), need to be carefully evaluated for each measure before an imputation method is selected.

To address data missingness, we first evaluate the frequency and patterns of missingness of each measures (e.g., missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR) [25, 26]), and then based on the nature of missingness, determine if and which imputation approach would be appropriate. For simple situation, we can use hot deck imputation, a non-parametric method of matching cases with missing values to observations with observed values for the same variable. However, for complex situations, hot deck imputation still has a potential for bias, a need for variance correction, and a need for having an adequate number of complete cases for matching. Thus, we need to use advanced missing data procedures such as the EM algorithm, fully Bayesian, maximum likelihood (ML) procedures, and the parametric multiple imputation (MI) methods, which are appropriate under general MAR conditions. ML and MI tend to yield similar results when implemented in comparable ways.

In order to reduce the possible impact of MNAR missingness for MACH14 longitudinal data, other known subject characteristics variables will be used in the missing data model with MI/ML, which can reduce estimation bias due to MNAR missingness, and partially restore lost power due to missingness [27, 28]. We will apply MNAR methods (e.g., selection models, pattern mixture models, and shared parameter models) to perform sensitivity analysis to evaluate model robustness [26].

## Conclusion

The MACH14 study has pooled and integrated existing data across multiple ARV adherence studies with compatible and similar study designs. It has enabled the building of a powerful and virtual large ART adherence study. The pooled data system with a large sample size and strong statistical power makes it possible to study questions that are otherwise difficult or impossible to address in a single small study. Sophisticated analytical methods that can deal with the inherent differences among studies are described and can be used by other studies.

## Acknowledgments

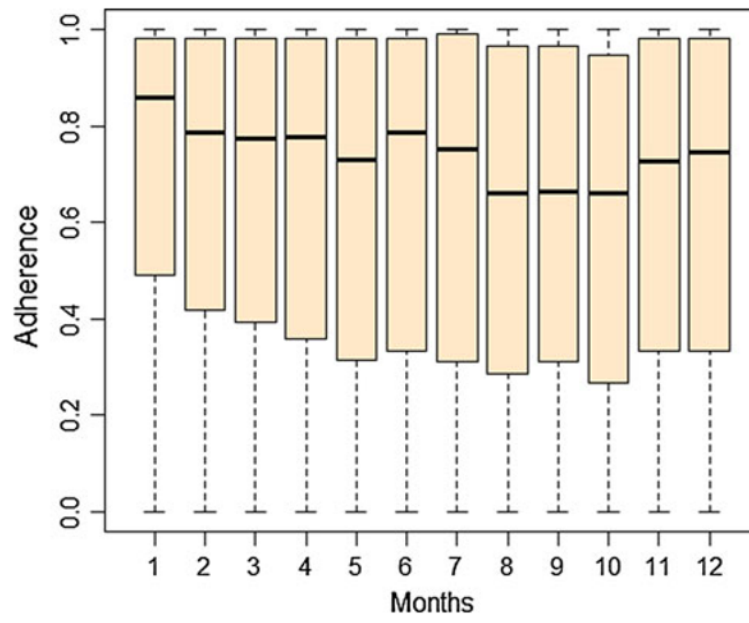
This research was supported by the multi-site adherence collaboration in HIV (MACH14) Grant R01MH078773 from the National Institute of Mental Health (NIMH), Office on AIDS. The original Grants of individual participating studies are: R01DA11869, R01MH54907, 2R01NR04749, R01NR04749, R01MH68197, R01DA13826, K23MH01862, R01MH01584, R01AI41413, R01MH61173, NIH/NIAIDAI38858, AI069419, K02DA017277, R01DA15215, NIMH P01MH49548, R01MH58986, R01MH61695, CC99-SD003, CC02-SD-003 and R01DA015679. We would like to thank all the subjects who participated in each of the individual studies. The

content of the paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

1. Feeney ER, Mallon PWG. HIV and HAART-associated dyslipidemia. *Open Cardiovasc Med J*. 2011; 5:49–63. [PubMed: 21643501]
2. Enanoria WTA, Ng C, Saha SR, Colford JM Jr. Treatment outcomes after highly active antiretroviral therapy: a meta-analysis of randomised controlled trials. *Lancet Infect Dis*. 2004; 4(7): 414–25. [PubMed: 15219552]
3. Bhaskaran K, Hamouda O, Sannes M, et al. Changes in the risk of death after HIV seroconversion compared with mortality in the general population. *JAMA*. 2008; 300(1):51–9. [PubMed: 18594040]
4. Bangsberg DR, Hecht FM, Charlebois ED, et al. Adherence to protease inhibitors, HIV-1 viral load, and development of drug resistance in an indigent population. *AIDS*. 2000; 14(4):357–66. [PubMed: 10770537]
5. Lucas GM. Antiretroviral adherence, drug resistance, viral fitness and HIV disease progression: a tangled web is woven. *J Anti-microb Chemother*. 2005; 55(4):413–6.
6. Chen LF, Hoy J, Lewin SR. Ten years of highly active antiretroviral therapy for HIV infection. *Med J Aust*. 2007; 186(3):146–51. [PubMed: 17309405]
7. Liu H, Golin CE, Miller LG, et al. A comparison study of multiple measures of adherence to HIV protease inhibitors. *Ann Intern Med*. 2001; 134(10):968–77. [PubMed: 11352698]
8. Reynolds NR. Adherence to antiretroviral therapies: state of the science. *Curr HIV Res*. 2004; 2(3): 207–14. [PubMed: 15279584]
9. Pearson CR, Simoni JM, Hoff P, Kurth AE, Martin DP. Assessing antiretroviral adherence via electronic drug monitoring and self-report: an examination of key methodological issues. *AIDS Behav*. 2007; 11(2):161–73. [PubMed: 16804749]
10. Wagner GJ. Predictors of antiretroviral adherence as measured by self-report, electronic monitoring, and medication diaries. *AIDS Patient Care STD*. 2002; 16(12):599–608.
11. Wagner G. Placebo practice trials: the best predictor of adherence readiness for HAART among drug users? *HIV Clinical Trials*. 2003; 4(4):269–81. [PubMed: 12916013]
12. Berg KM, Arnsten JH. Practical and conceptual challenges in measuring antiretroviral adherence. *J Acquir Immune Defic Syndr*. 2006; 43(Suppl 1):S79–87. [PubMed: 17133207]
13. Levine AJ, Hinkin CH, Marion S, et al. Adherence to antiretroviral medications in HIV: differences in data collected via self-report and electronic monitoring. *Health Psychol*. 2006; 25(3):329–35. [PubMed: 16719604]
14. Nachega JB, Hislop M, Dowdy DW, et al. Efavirenz versus nevirapine-based initial treatment of HIV infection: clinical and virological outcomes in Southern African adults. *AIDS*. 2008; 22(16): 2117–25. [PubMed: 18832875]
15. Rivet Amico K, Harman JJ, Johnson BT. Efficacy of antiretroviral therapy adherence interventions: a research synthesis of trials, 1996 to 2004. *JAIDS J Acq Immune Defic Syndr*. 2006; 41(3):285–297.
16. Konkle-Parker DJ, Erlen JA, Dubbert PM. Lessons learned from an HIV adherence pilot study in the Deep South. *Patient Educ Couns*. 2010; 78(1):91–6. [PubMed: 19560307]
17. Giordano TP, Guzman D, Clark R, Charlebois ED, Bangsberg DR. Measuring adherence to antiretroviral therapy in a diverse population using a visual analogue scale. *HIV Clin Trials*. 2004; 5(2):74–9. [PubMed: 15116282]
18. Pinheiro, JC.; Bates, DM. *Mixed-effects models in S and S-PLUS*. New York: Springer; 2009.
19. Teramukai S, Matsuyama Y, Mizuno S, Sakamoto J. Individual patient-level and study-level meta-analysis for investigating modifiers of treatment effect. *Jpn J Clin Oncol*. 2004; 34(12):717–21. [PubMed: 15640501]
20. Jones AP, Riley RD, Williamson PR, Whitehead A. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clin Trials*. 2009; 6(1):16. [PubMed: 19254930]

21. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol.* 2002; 55(1):86–94. [PubMed: 11781126]
22. Cooper H, Patall EA. The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychol Methods.* 2009; 14(2):165–76. [PubMed: 19485627]
23. Liu H, Miller LG, Hays RD, et al. A practical method to calibrate self-reported adherence to antiretroviral therapy. *JAIDS J Acq Immune Defic Syndr.* 2006; 43:S104–12.
24. Liu H, Miller LG, Hays RD, et al. Repeated measures longitudinal analyses of HIV virologic response as a function of percent adherence, dose timing, genotypic sensitivity, and other factors. *JAIDS J Acq Immune Defic Syndr.* 2006; 41(3):315–22.
25. Little, RJA.; Rubin, DB. *Statistical analysis with missing data.* 2. New York: Wiley-Interscience; 2002.
26. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods.* 2002; 7(2):147–77. [PubMed: 12090408]
27. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods.* 2001; 6(4):330–51. [PubMed: 11778676]
28. Newgard CD, Haukoos JS. Advanced statistics: missing data in clinical research—part 2: multiple imputation. *Acad Emerg Med.* 2007; 14(7):669–78. [PubMed: 17595237]



**Fig. 1.**  
Overall adherence of 16 studies by month

Table 1

## General information about MACH14 16 participating studies

Institute	Study name	Interventional/ observational	Description of intervention	Type	Length of intervention	Project dates	No. of subjects	Length of follow-up
University of Miami	Behavioral management and stress responses in HIV/AIDS	Interventional	Randomized design, about 50% ( $n = 227$ ) cognitive-behavioral stress management (CBSM) group-based psychosocial intervention, 50% ( $n = 170$ ) regular control. Participants received medication adherence training by a clinical pharmacist	Psychosocial intervention	10 weeks	1997–2003	404	18 months
University of Pittsburgh	Improving adherence to antiretroviral therapy	Interventional	Subjects are randomized to one of three groups of structured and individualized interventions and usual care (control) with the 100% adherers being followed separately. Both interventions are being delivered by telephone. Subjects in the intervention groups are further randomized to booster/no booster following the maintenance program. Intervention was based on social cognitive theory and self efficacy theory	Telephone interventions: structured or individualized. Tapered maintenance program.	28 weeks	2003–2009	349	19 months
University of Washington	Peer and pager support to enhance antiretroviral adherence (PAL)	Interventional	About 180 subjects are in the intervention. The pager arm has 60, the “buddy” arm has 60, and another 60 received both “buddy” and pager. The intervention is based on social support theory	Pager; “buddy”; both “buddy” and pager	12 weeks	2002–2009	224	9 months
University of Pittsburgh	Adherence to protease inhibitors	Interventional	Subjects were randomized to treatment ( $n = 99$ ) or usual care/control ( $n = 101$ ). 15 perfect adherers (100% adherence) were followed separately. A 12 week structured telephone intervention was provided, followed by a 16 week tapered maintenance program. Intervention was based on social cognitive theory and self efficacy theory	Structured telephone intervention; tapered maintenance program	28 weeks	1998–2003	215	13 months
Columbia University & NYSPH	Serodiscordant couples, medical adherence and HIV risk (SMART)	Interventional	106 were randomized to the intervention arm. The intervention consisted of four 45–60 min sessions delivered over 5 weeks (between the Baseline and Week 8 assessment) by a Nurse Practitioner in a clinic setting. Intervention was grounded in Ewart’s Social Action Theory and aimed to improve adherence.	Counseling	5 weeks	2000–2004	215	8 months
University of Missouri-Kansas City	Couples study ART adherence: enhanced counseling and observed therapy	Interventional	Skills, motivation, & support for adherence from relationship partner Three arms ( $n = 202$ total). 2 of 3 groups (137 subjects) receive adherence intervention [Motivational interviewing (MI) counseling alone (69 subjects), or MI plus modified observed therapy with one dose observed every day (68 subjects)] and the third group (65 subjects) was standard care. The intervention was based on information-motivation-behavioral skills, Motivational Interviewing and self-determination theory	MI counseling alone; MI plus modified observed therapy	24 weeks	2004–2008	202	12 months
RAND	California cooperative treatment group (CCTG) 578	Interventional	The intervention consisted of adherence counseling using cognitive-behavioral components administered by research nurses over 5 week sessions (3 prior to starting ART, and two during first 2 weeks of ART); there were 2 intervention arms in the study—one received the cognitive behavioral intervention (66 subjects, and the other received the cognitive behavioral intervention plus a 2-week pre-ART practice trial with placebo pills (66 subjects). And a non-intervention control (67 subjects) uses social cognitive theory	Cognitive behavioral intervention or cognitive-behavioral intervention plus 2-week pre-ART practice trial;	5 weeks	2000–2002	199	12 months
Tufts University	Understanding and improving adherence in HIV disease	Interventional	The primary intervention ( $n = 111$ ) consists of an ARV adherence report shared with physicians. This report summarizes the subject’s adherence between visits using both self-reported items and electronic MEMS data	Randomized, report sharing, cross-over, intervention study	20 weeks	2001–2003	156	24 months
University of North Carolina (UNC) at Chapel Hill	Participating and communicating together (PACT)	Interventional	2-arm block, randomized, controlled design to compare adherence (at 12-week follow-up) of subjects receiving a Motivational Interviewing (MI) intervention with those receiving a dose-matched HIV informational control program. Intervention consisted of 3 components: (1) a 20-minute audiotape and booklet immediately before seeing their medical provider; (2) 2 one-on-one sessions with a health educator at weeks 4 and 8 of follow-up, and (3) a mailing 2 weeks after each individual session. Control group materials were comparable to the intervention in length and format to control for the placebo effect of exposure but provided general HIV information only	MI counseling	12 weeks	1999–2004	155	3 months
University of California, Los Angeles (UCLA)	Adherence and efficacy of protease inhibitor therapy (ADEPT)	Observational	N/A	N/A	N/A	1997–2000	145	12 months
Ohio State University	AIDS clinical trial group (ACTG) 751	Interventional	2-arm study. Subjects were randomly assigned to receive either standard ACTG clinic-based subject education (SC = 55) or, in addition to standard care, a more intensive approach (TX = 53) that included telephone calls delivered at regular intervals by a trained, registered nurse over the first 16 weeks of antiretroviral therapy. Structured phone calls, conducted	Structured telephone support	16 weeks	1998–2003	109	16 months

Institute	Study name	Interventional/ observational	Description of intervention	Type	Length of intervention	Project dates	No. of subjects	Length of follow-up
University of California, San Francisco (UCSF)	Research in access to care in the homeless (REACH)	Observational	N/A	N/A	N/A	1997–2002	108	60 months
Albert Einstein College	HIV epidemiology research on outcomes (HERO adherence study)	Observational	N/A	N/A	N/A	1998–2004	104	6 months
University of North Carolina (UNC) at Chapel Hill	Directly observed therapy (DOT)	Interventional	Directly observed therapy (DOT) (25 subjects) and regular kept on person (KOP) control (75 subjects)	Directly observed therapy	48 weeks	2000–2005	102	12 months
Yale University	Rewards improve medication compliance for HIV treatment (REWARDS)	Interventional	Randomized controlled trial comparing supportive counseling to counseling involving review of MEMS data and prizes for MEMS-verified medication-taking. Participants used MEMS caps for 4 weeks. Those with adherence to dose-time less than 80% were randomly assigned to 16 weeks counseling and 16 weeks follow-up. (16 weeks of weekly meetings.) Based on contingency management theory	Supportive counseling vs. counseling with MEMS review and reinforcement.	16 weeks	2002–2005	97	9 months
University of Pennsylvania	Adherence to protease inhibitors in HIV	Observational	N/A	N/A	N/A	2005–2006	76	4 months

from a central site, were tailored to the individual and were designed to address common barriers to adherence and promote self-care strategies. The intervention was guided by self-regulation theory

**Table 2**  
**General statistics of MACH14 demographic data, MEMS data, viral load data, CD4 data and drug resistance data for all 16 cohorts**

Study institute	No. of subjects	MEMS		Viral load		CD4		Drug resistance	
		No. of subjects	No. of readings	No. of subjects	No. of readings	No. of subjects	No. of readings	No. of subjects	No. of readings
University of Miami	404	258	63,322	397	1,198	397	1,198	—	—
University of Pittsburgh	349	345	18,717	336	1,117	336	1,114	—	—
University of Washington	224	195	33,150	219	753	222	764	—	—
Columbia University & NYSP	215	211	13,733	214	732	213	745	—	—
University of Pittsburgh	215	205	13,697	208	616	209	594	—	—
University of Missouri-Kansas City	202	197	51,046	202	764	201	742	59	348
RAND	199	185	16,244	198	815	198	797	—	—
Tufts University	156	155	49,062	156	767	155	700	—	—
UNC at Chapel Hill	155	84	6,763	150	259	—	—	—	—
UCLA	145	123	47,908	138	1,067	92	215	55	297
Ohio State University	109	105	22,968	109	2,041	109	1,745	—	—
UCSF	108	104	26,771	103	4,046	103	2,912	65	692
Albert Einstein College	104	81	21,974	104	672	104	104	—	—
UNC at Chapel Hill	102	99	47,641	102	1,001	102	698	19	96
Yale University	97	97	34,735	85	151	79	79	38	98
University of Pennsylvania	76	54	10,511	55	251	55	125	—	—
Total	2,860	2,498	478,242	2,776	16,250	2,575	12,532	236	1,531

Note: “—” indicates the data are not available



**Table 3**  
**Demographic characteristics of MACH14 sample (N = 2,860)**

Characteristics	Sample frequency	N (%) with available data <sup>a</sup>
Age mean (SD)	41 (8.3)	2,835 (99.1)
Gender, <i>n</i> (%)		2,837 (99.2)
Male	1,929 (68.0)	
Female	908 (32.0)	
Education, <i>n</i> (%)		2,730 (95.5)
Less than high school	663 (24.3)	
High school diploma/GED	1,715 (62.8)	
More than high school	352 (12.9)	
Race, <i>n</i> (%)		2,789 (97.5)
Black or African-American	1,337 (47.9)	
Caucasian or White	800 (28.7)	
Hispanic or Latino	409 (14.7)	
Asian American/Native American/Multiracial	243 (8.7)	
Employment, <i>n</i> (%)		2,507 (87.7)
Unemployed	1,738 (69.3)	
Employed (part-time/full-time)	769 (30.7)	
Sexual orientation, <i>n</i> (%)		1,710 (59.8)
Heterosexual	700 (40.9)	
Homosexual	581 (34.0)	
Bisexual	246 (14.4)	
Not sure	183 (10.7)	
Born, <i>n</i> (%)		1,080 (37.8)
US	901 (83.4)	
Other	179 (16.6)	
Jail, <i>n</i> (%)		795 (27.8)
Yes	348 (43.8)	
No	447 (56.2)	
NAÏVE, <i>n</i> (%)		2,675 (93.5)
Naïve to antiretroviral medications at baseline	476 (17.8)	
Not Naïve	2,199 (82.2)	
Men who have sex with men, <i>n</i> (%)		2,302 (80.5)
Yes	1,004 (43.6)	
No	684 (29.7)	
N/A (for female)	614 (25.6)	
Injection drug use, <i>n</i> (%)		1,907 (66.7)
Yes	357 (18.7)	
No	1,550 (81.3)	
Other exposure, <i>n</i> (%)		1,728 (60.4)
Yes	689 (39.9)	

Characteristics	Sample frequency	<i>N</i> (%) with available data <sup>a</sup>
No	1,039 (60.1)	
Income, <i>n</i> (%)		2,317 (81.0)
Less than 10 k	1,392 (58.2)	
10–20 k	494 (20.7)	
20–30 k	178 (7.4)	
30–40 k	112 (4.7)	
40 k	141 (5.9)	
Refused	49 (2.0)	
Housing <sup>b</sup> , <i>n</i> (%)		1,672 (58.5)
Rent/own	842 (50.3)	
A friend's/relative's home	275 (16.4)	
Transitional/subsidized house	93 (5.6)	
Homeless shelter	23 (1.4)	
Homeless	259 (15.5)	
Residential drug facility	84 (5.0)	
Hospitalized	1 (0.1)	
Nursing home/personal care	0 (0.0)	
Other	95 (5.7)	
Marriage status, <i>n</i> (%)		1,815 (63.5)
Currently in a committed relationship	498 (27.4)	
Not in a committed relationship	1,288 (71.0)	
Refused or other	29 (1.6)	

<sup>a</sup>*N* refers to the number of people with valid values, and percentage refers to the percentage of *N* relative to the total sample size (*N* = 2,860)

<sup>b</sup>Housing is combined from “homeless (Yes/No)” and “housing (9 categories)”, by adding all of the homeless from the “homeless” variable to the homeless in the “housing” variable and then adding the remainder to the “Other” category of the “housing” variable

Table 4

Antiretroviral (ARV) characteristics grouped by ARV type

ARV code	Name of ARV	No. of readings	No. of subjects	Person months	ARV type	No. of drugs
ENF	Enfuvirtide (Fuzeon, ENF, T-20)	499	5	4.07	FI	1
EFV	Efavirenz (Sustiva)	37,537	290	5.89	NNRTI	1
NVP	Nevirapine (Viramune)	29,833	156	4.95	NNRTI	1
ATR	Atripla (efavirenz + TDF+FTC)	5,943	27	9.37	NNRTI	3
DLV	Delavirdine (Rescriptor)	5,457	42	4.29	NNRTI	1
CBV	Combivir (AZT + 3TC)	37,424	240	3.82	NRTI	2
TRZ	Trizivir (AZT+3TC + ABC)	20,924	128	4.01	NRTI	3
ABC	Abacavir (Ziagen)	16,233	106	3.79	NRTI	1
D4T	Stavudine (Zerit)	14,260	103	3.61	NRTI	1
3TC	Lamivudine (Epivir)	13,895	126	3.48	NRTI	1
DDI	Didanosine (Videx)	8,101	81	3.37	NRTI	1
AZT	Zidovudine (Retrovir, AZT, ZDV)	5,758	37	3.66	NRTI	1
ADF	Adefovir (ADF)	1,363	20	3.37	NRTI	1
DDC	Zalcitabine (Hivid)	949	9	3.17	NRTI	1
EPZ	Epizicom (ABC + 3TC)	713	17	2.68	NRTI	2
TDF	Tenofovir (Viread)	3,656	48	3.46	NRTI	1
TRU	Truvada (FTC + TDF)	2,969	48	3.11	NRTI	2
FTC	Emtricitabine (Emtriva)	2,103	23	3.87	NRTI	1
NFV	Nelfinavir (Viracept, NFV)	77,244	406	4.62	PI	1
LPV	Lopinavir/Ritonavir (Kaletra)	58,128	356	4.47	PI	2
IDV	Indinavir (Crixivan, IDV)	36,240	171	4.15	PI	1
ATZ	Atazanavir (Reyataz)	28,276	215	6.39	PI	1
RTV	Ritonavir (Norvir, RTV)	20,506	148	5.51	PI	1
SQV	Saquinavir (Fortovase, SQV)	16,697	88	6.04	PI	1
FOR	Saquinavir (Invirase, SQV)	7,509	71	3.36	PI	1
FPV	Fosamprenavir (Lexiva, Telzir, FPV)	6,629	44	6.31	PI	1
APV	Amprenavir (Agenerase)	2,522	30	2.88	PI	1
DRV	Darunavir (Prezista)	1,870	6	9.00	PI	1
TPV	Tipranavir (Aptivus, TPV)	1,247	5	7.32	PI	1

ARV code	Name of ARV	No. of readings	No. of subjects	Person months	ARV type	No. of drugs
PLC	Placebo	12,266	54	6.20	-	-
OTH	Other drugs	1,236	28	1.55	-	-
Total		477,987			-	-

Note: “-” indicates the data are not available

Table 5

## Viral Load and CD4 distributions of MACH14 sample

First recorded viral load	Frequency	Percent
Undetectable (<50 copies/ml)	585	22.4
Detectable (≥ 50 copies/ml)	2,031	77.6
<i>N</i>	Mean	SD
	Minimum	25th Percentile
	Median	75th Percentile
	Maximum	
First recorded viral load (log <sub>10</sub> ) <sup>a</sup>		
2,616	3.32	1.58
	0	1.88
	3.25	4.71
	6.81	
Viral load per subject (log <sub>10</sub> ) <sup>a</sup>		
2,676	2.87	1.27
	0	1.89
	2.78	3.78
	6.57	
First recorded CD4		
2,564	359	286
	0	155
	299	492
	3,029	
CD4 per subject		
2,572	396	397
	1	190
	344	525
	15,070	

<sup>a</sup>Viral load values are log<sub>10</sub> based, and to make it non-negative, 1 is added before taking logarithms

**Table 6**Clinical data distributions of MACH14 sample ( $N = 2,860$ )

Characteristics	Sample frequency	$N$ (%) with available data <sup>a</sup>
Lowest ever absolute CD4 count (cells/mm <sup>3</sup> ), mean (SD)	213 (212.9)	1,207 (42.2)
Duration of known HIV status at time of enrollment (month), mean (SD)	15.9 (31)	2,118 (74.1)
Highest viral load in record (log <sub>10</sub> copies/ml), mean (SD)	4.05 (0.7)	2,370 (82.9)
CDC stage at time of entry into study, $n$ (%)		2,059 (72.0)
A	174 (8.5)	
B	114 (5.5)	
C	320 (15.5)	
Unknown	1,451 (70.5)	
Currently use or ever used any kind of alternative or complementary medications, treatments, or supplements, $n$ (%)		2,274 (79.5)
No	687 (30.2)	
Yes	739 (32.5)	
Unknown	848 (37.3)	
Ever had an HIV-related infection or malignancy or complication, $n$ (%)		2,274 (79.5)
No	601 (26.4)	
Yes	804 (35.4)	
Unknown	869 (38.2)	
Ever taken or currently on PCP prophylaxis, $n$ (%)		2,232 (78.0)
No	281 (12.6)	
Yes, at enrollment	159 (7.1)	
Yes, only for a time before enrollment	12 (0.5)	
Yes, only at a time starting after enrollment	62 (2.8)	
Yes, for a time prior to enrollment and after enrollment	57 (2.6)	
Yes during the study, but unknown if prior to study	142 (6.4)	
Known	1,519 (68.1)	
Ever taken or currently on TB prophylaxis, $n$ (%)		2,017 (70.5)
No	156 (7.7)	
Yes, at enrollment	27 (1.3)	
Yes, only for a time before enrollment	22 (1.1)	
Yes, only at a time starting after enrollment	32 (1.6)	
Yes, for a time prior to enrollment and after enrollment	15 (0.7)	
Yes during the study, but unknown if prior to study	7 (0.3)	
Unknown	1,758 (87.2)	

<sup>a</sup> $N$  refers to the number of people with valid values, and the percentage is the percentage of the total sample ( $N = 2,860$ )