*CONFIDENTIAL*

# IMPAACT 2002

# Combined Cognitive Behavioral Therapy and a Medication Management Algorithm for Treatment of Depression among Youth Living with HIV in the United States

## Study Analysis Plan v3.0

## September 27, 2019

## Protocol version 1.0

## LOA #1

## CT.gov NCT02939131

**This is IMPAACT 2002 SAP Version 3.0 with names of authors, names of publication writing team members and analysis timeline redacted; these edits were done on July 14, 2020.**

***Filename:*** S:/2002/Documentation/analysis_plan/IMPAACT 2002 SAP v3.0 20190927 for CT.gov.docx
***Version:*** *3.0 edited for CT.gov submission*
***Directory mapping:*** *S:/Disney/home/impaact*
***Printed on:*** *July 14, 2020*

## Table of Contents

## Table of Tables

## 1 Introduction

This document describes the general analytical approach for the primary objectives (week 24) of IMPAACT 2002, indicating which analyses will be primary and which will be supplementary. In addition, we outline approaches for the secondary and exploratory objectives and clarify the timing of these analyses. This plan also outlines our approach for analyzing interim data to check on statistical assumptions behind the power analysis, an analysis which was completed in the fall of 2018. It is recognized that the analysis plan may be modified by the Study Team to reflect recommendations made by the Study Monitoring Committee (SMC). The primary (week 24) analysis will be carried out while the study is still in follow-up. Timetables for the Week 24 and Week 48 analyses are shown in Appendix 1.

## 2 Version history

| Version | Changes made | Date finalized |
|---|---|---|
| 1 | Original | 7/21/2017 |
| 2.0 | Updated personnel, Updated text relating to LOA#1; Updated text relevant to protocol history, SMC reviews and interim monitoring findings. | 3/5/2019 |
| 3.0 | Revised Section 7 to more clearly delineate primary and supplementary statistical analyses. Also clarified that the interim analysis has been completed, accrual has been completed and that an additional SMC review was also completed. Refined timeline for primary and secondary analyses to correspond to MOP, Section 19. | 9/27/2019 |
| 3.0 | Prepared for CT.gov. Redacted authors on title page, Study team members and Appendices with analysis timelines and LOA approval dates. | 7/14/2020 |

# 3 Core Protocol Team (Redacted)

Redacted on July 14, 2020.

# 4 Study Overview

## 4.1 Study Design

IMPAACT 2002 study is a prospective, multi-site, two-arm cluster (site)-randomized study of the efficacy of a health and wellness cognitive-behavioral intervention (H&W CBT) and medication management (COMB-R) for depression in youth with HIV compared to enhanced standard care (ESC). The study population are HIV-infected youth in the U.S., who are between 12 and 24 years old (inclusive), and who have been diagnosed with moderate to severe depression reflected by a QIDS-C score of ≥11. Youth enrolled in the study attend a Screening/Entry Visit and study visits at Weeks 1, 6, 12, and 24. Participants are encouraged to attend all clinically indicated interim visits through the first 24 study weeks. Two additional follow-up visits are scheduled at Weeks 36 and 48. The primary objectives are to compare the two study arms with respect to depression outcomes, and biological measures of health after 24 weeks.

Fourteen identified U.S. IMPAACT sites were randomly assigned to either the COMB-R group or the ESC group. Randomization of sites was chosen due to the potential contamination that could occur if same site clinicians were delivering both ESC and COMB-R. A pre-study survey ascertained broadly defined site characteristics, namely, the number of youth served and whether youth were primarily behaviorally or perinatally infected with HIV. A follow-up pre-randomization survey queried sites for more specific characteristics of potential study participant, which included gender, age (12-18 versus 19-24 years of age), viral suppression status (below versus not below the level of detection as defined by the laboratory), transmission routes (perinatal versus behavioral HIV acquisition), and initial levels of depression (moderate vs. severe). A restricted randomization procedure was used to assign sites to the two study arms in a way that balanced these characteristics across arms. However, after randomization, one site assigned to the COMB-R treatment arm withdrew from the study. Thus the final randomization included seven sites assigned to ESC and six to COMB-R, for a total of 13 sites.

Initially the protocol followed procedures to ensure representativeness of clinic populations and to reduce selection bias. Specifically, after sites were registered, they submitted lists of potential study participants by means of a screening log. These potential study participants were then randomly ordered within each site in blocks of six. Sites were instructed to approach the study participants in the block order given and

not go to a new block of participants until all participants in the prior block had been approached and a decision reached as to study participation. Participants within a specific block could be approached in any order. Sites could continue to periodically submit new lists of potential participants, which included re-approaching potential participants who initially had failed to enroll. A screening failures form was completed for each study participant who was approached but did not enroll. Data on a limited number of characteristics (gender, age group, viral suppression status, transmission routes, and initial levels of depression) were gathered for those participants who failed to enroll, but only if they had initially consented to enroll. In addition, reasons for screening failures were collected for those participants who had initially consented. .

A letter of amendment (LOA#1), dated April 27, 2018, removed the requirement to follow a predefined random ordering for approaching participants. After LOA#1 implementation, sites could approach any participant in their clinic and request a screening number. Once approached, if the participant chose not to enroll, a screening failures form was completed as noted above. Appendix 2 indicates the date at which time the LOA#1 was approved at each site and the requirement for block-random ordering approach removed. Resolution of screening numbers for participants who had been block-randomized but not yet approached at the time the site implemented LOA#1 is described below.

### 4.2 Hypotheses

- Participants receiving COMB-R will demonstrate improved depression outcomes (e.g., decreased depressive symptoms, and greater remission and response rates) compared to participants receiving ESC.
- Participants receiving COMB-R will demonstrate improved medical outcomes (e.g., increased CD4 T-cell count, decreased HIV RNA level) compared to participants receiving ESC.

### 4.3 Primary Objectives and Outcomes

1. To evaluate whether the Health and Wellness Cognitive Behavioral Therapy and Medication Management Algorithm (COMB-R) treatment for depression is associated with improved depression outcomes at 24 weeks, compared to Enhanced Standard Care (ESC).
   Outcomes:
   - QIDS-SR score at Week 24 (primary efficacy outcome)
   - Response to Treatment, defined as a decrease in QIDS-SR score by >50% from entry to week 24
   - Remission, defined as a QIDS-SR score $\leq$ 5 at week 24
2. To evaluate whether the COMB-R is associated with improved biological measures of health over 24 weeks, including CD4 cell numbers and copies of HIV RNA in plasma compared to ESC.
   Outcomes:
   - CD4 cell count at Week 24

- Plasma HIV RNA level at Week 24

### 4.4 Secondary Objectives and Outcomes

Except as noted in Section 7.4 below, the secondary study objectives will be analyzed after the week 48 data are complete, approximately six months after the primary (week 24) data analyses.

1. To assess whether COMB-R is associated with improved adherence for HIV and depression treatment compared to ESC for the first 24 weeks and whether any differences are maintained at 48 weeks.
   Outcomes:
   - Self-reported adherence to anti-HIV medications: : # days of 30 with missed doses, how good participant was at taking medicines as instructed, how often did participant take medicines as instructed
   - Self-reported adherence to psychiatric medications: : # days of 30 with missed doses, how good participant was at taking medicines as instructed, how often did participant take medicines as instructed
   - Adherence to study visits: # visits completed
   - Adherence to psychotherapy sessions: # sessions attended
   - Adherence to medication management visits (COMB-R): # sessions attended

2. To assess whether differences in depression treatment outcomes are maintained at 48 weeks.
   Outcomes:
   - QIDS-SR score at Week 48 (primary efficacy outcome)
   - Response to Treatment, defined as a decrease in QIDS-SR score by >50% from entry to week 48
   - Remission, defined as a QIDS-SR score $\leq$ 5 at week 48
3. To assess whether demographic, behavioral, and biological factors could moderate the efficacy of COMB-R compared to ESC arms
   - Demographic: age, sex at birth;
   - Behavioral: HIV acquisition category, initial level of depression;
   - Biological: baseline CD4, nadir CD4, plasma HIV RNA, CDC category
4. To assess whether COMB-R is associated with improved behavioral risk outcomes (alcohol/drug use; sex-risk behaviors) compared to those on ESC at Week 24 and Week 48
   Outcomes
   - Alcohol use: ever used, past 3 months frequency, # drinks/day, binge drinking
   - Tobacco use: ever used, past 3 months frequency

- Drug use: ever used, past 3 months frequency for: cannabis, cocaine, amphetamine, inhalants; sedatives, hallucinogens, opioids.
- Sex-risk behaviors: Condom use (importance, confidence, frequency), sex as exchange commodity, number of sex partners

5. To describe the implementation fidelity at COMB-R sites and the counseling strategies and medication patterns at ESC sites.

   Outcomes

   - COMB-R: MM fidelity: # sessions administered and stages (Table 1);
   - COMB-R: CBT fidelity: # sessions administered; CBT items covered (Table 2)
   - ESC: counseling fidelity: # sessions administered; Counseling approaches (Table 3)

6. To describe and compare the number of interim visits, the frequency of medication use, and acceptability of COMB-R and ESC among participants and clinicians

   Outcomes

   - Number of interim visits (with site prescribing clinician; with counseling clinician) scheduled and attended during first 24 weeks
   - Frequency of medication use during first 24 weeks: percent of time on medications (by stage; Table 1)
   - Patient acceptability during first 24 weeks
     - Client satisfaction (sum or mean of 8 items)
   - Clinician acceptability during first 24 weeks
     - ESC prescribing clinician satisfaction
     - ESC counseling clinician satisfaction
     - COMB-R prescribing clinician satisfaction
     - COMB-R CBT clinician satisfaction

7. To compare the frequency and types of new (post-baseline) grade 3 or higher Adverse Events, psychological hospitalizations and suicide attempts between COMB-R and ESC arms

   Outcomes

   - New (post-baseline) grade 3 or higher Adverse Events
   - Psychological hospitalizations
   - Suicide attempts

### 4.5 Exploratory Objectives

1. To examine whether COMB-R is associated with a larger decrease in plasma inflammatory biomarkers from entry to weeks 24 and 48 compared to ESC
2. To evaluate the moderating effect of inflammatory markers on the efficacy of COMB-R compared to ESC.

### 4.6 Key Eligibility Criteria

Eligibility criteria include being 12-24 years (inclusive), having documented HIV infection and being aware of their infection status. Participants have a primary diagnosis of nonpsychotic depression as defined by DSM-IV or DSM-V criteria and their current depressive symptoms warrant intervention as determined by a score of 11 or greater on the clinician version of the Quick inventory of Depressive Symptomatology (QIDS-C). Participants must be able to communicate in spoken and written English.

Key exclusionary criteria include having severe or moderate alcohol or substance use problems, having depression or suicidal ideation that may require more intensive treatment than the study provides and not wishing to switch providers to one trained on study material.

### 4.7 Visit and Evaluation Schedule

Participants at every COMB-R and ESC assigned site will attend a Screening/Entry Visit, and follow-up clinic visits at Weeks 1, 6, 12, 24, 36, and 48. In addition, participants in the COMB-R sites will be encouraged to return to the clinic for weekly therapy visits through Week 8, every other week through Week 16, and then monthly until Week 24. These visits are not mandatory, and will be considered and documented as interim visits. At these interim visits, study staff will conduct therapy utilizing the CBT manual and provide medication management utilizing the MM manual, as indicated. Participants in the ESC sites will likewise return to the clinic for interim visits, based on the clinical indication of the participant.

### 4.8 Protocol History

Protocol Version 1.0 was dated August 19, 2016. Randomization was completed November 22, 2016. The protocol opened to accrual on December 20, 2016. A clarification memo was finalized March 6, 2017 in order to clarify study recruitment, screening and enrollment procedures, clarify Clinician Satisfaction Scale forms and update the Protocol Team and Site Roster.

A second clarification memo dated October 4, 2017 clarified completion of the off study form, that both ACASI and paper QIDS-SR questionnaires must be completed if the Week 1 study visit occurred on the same day as the Screen/Entry visit and to specify use of Version 2.1 of the Division of AIDS Table for Grading the Severity of Adult and Pediatric Adverse Events. The protocol team roster was also updated.

As noted above, a letter of amendment (LOA#1) was implemented on April 27, 2018, which removed the requirement to approach participants in a pre-defined block-randomized order once sites received IRB approval.

Prior to enrolling participants, each site had to obtain a screening number for each potential participant. At each site, prior to approval of LOA#1, these screening numbers were then randomly ordered in blocks of

size six and the sites were instructed to approach the potential participants in the block order indicated. Before moving onto potential participants in a second block, all participants in the prior block had to be approached. If a participant chose not to enroll or could not be reached, a screening failure form was completed, which, if consent was given, collected selected data on personal characteristics and reasons for eligibility failure. The first set of randomly ordered screening numbers was sent to the sites on February 21, 2017 and the first participant was enrolled on March 6, 2017.

After approval for LOA#1, the requirement for site-specific block-randomization was removed. Dates of implementation of LOA#1 by site are listed in Appendix 2. Screening numbers for participants who were block-randomized but had not yet been approached at the implementation of LOA#1 were resolved with a screening failure form indicating refusal of consent. In order to distinguish these screening failures from those participants who were actually approached but either refused consent or failed screening, data managers also entered a comment field corresponding to each screening number, with an indication that resolution was due to the implementation of LOA#1.

### 4.9 Monitoring

Study data and conduct are monitored through a combination of reports, including monthly accrual reports and quarterly monitoring reports, which include data on imbalance in characteristics between study arms, baseline characteristics, study status, visit and form completion, data availability and treatment implementation fidelity. Safety (diagnoses, new post-baseline grade 3 and higher adverse events, psychological hospitalizations and suicide attempts) will also be monitored. Study conduct and safety are also monitored through annual Study Monitoring Committee (SMC) reviews. If either safety trigger (defined below) is met, the SMC determines whether to convene an *ad hoc* review of participant safety. In addition to these reviews, in order to verify assumptions underlying the study design, an interim analysis of the intraclass correlation of the primary study outcome (QIDS-SR ACASI week 0 score) was carried out as described below.

#### 4.9.1 Accrual

Accrual is monitored monthly. It was expected that full accrual to the study would take two years. At the time of this plan revision, the study has completed the target accrual (N=156) and done so within approximately two years of the date the first participant was accrued: the first and last participants were enrolled in March 2017 and March 2019, respectively.

#### 4.9.2 Randomization properties: Monitoring for imbalance between study arms

During enrollment, the core protocol team monitors the balance of the selected characteristics between arms quarterly or as needed (blinded to which arm is which), and if an imbalance greater than the maximum allowable imbalance is identified, may impose enrollment limits in the DMC enrollment system for all sites in one arm or the other to force subsequent enrollments to be from a specific category of age

(<18 years or > 18 years), gender (male or female), HIV acquisition route (perinatal or behavioral), depression level (moderate or severe) or viral suppression (suppressed or not suppressed), as needed, to achieve approximate balance of these characteristics in the two study arms. See next section on SMC History for a review of imbalance criterion.

### 4.10 SMC History

The study had an initial SMC review on September 30, 2016 prior to study opening to accrual. It was suggested we include safety triggers in the monitoring plan and also track screening failure reasons for potential causes of concern at individual sites. A second SMC review was performed on July 14, 2017, with the primary goal of reviewing accrual and randomization. Subsequent SMC reviews were performed on December 15, 2017; June 1, 2018, December 21, 2018 and in late June 2019. Also see below for the SMC's review of safety data.

#### 4.10.1 Review of imbalance criterion

The imbalance criterion was initially set to a group difference of 25%. After the SMC review of June 2018, due to concerns about the imbalance in level of depression nearing the criterion, the study team was asked to review how imbalance in depression was assessed and to develop a plan for further monitoring. The study team thereafter proposed evaluating depression levels based on the structured QIDS-C (Clinician) report at screening/entry rather than the unstructured judgement recorded on the eligibility checklist. At the same time, the study team proposed reducing the criterion for imbalance to 20%. As of the date of this report, balance has been maintained.

## 5 Safety

As noted above in Section 4.9, safety endpoints (significant diagnoses, new post-baseline grade 3 or higher adverse events, psychiatric hospitalizations and suicide attempts) are monitored quarterly and if safety triggers, as defined below, are met, the SMC can decide whether to convene additional reviews.

### 5.1 Safety Triggers

Either of the following constitutes a safety trigger:

- Percent of participants (of those enrolled) who are hospitalized for a psychiatric indication or who attempt suicide exceeds 5%;;
- Two or more participants at a given site are hospitalized for a psychiatric indication or attempt suicide.

Through routine monitoring, the Protocol Team determined that the safety trigger had been met in May 2018. The SMC reviewed these data and was kept informed of additional safety events contributing to the assessment of the safety trigger.

# 6 Interim Analysis

An interim analysis was planned either (a) once at least five participants at each site have completed Week 0 or (b) 18 months from the start of study accrual, whichever occurred first. The purpose of the interim analysis was to estimate the magnitude of the intracluster correlation (ICC) of the QIDS-SR ACASI score in order to validate the sample size computations. If this analysis showed that the estimated ICC was larger than 0.15 (the largest value for which the selected sample size provides 80% power), the protocol team would assess whether to increase the number of participants to be enrolled, to ensure adequate statistical power for the primary study objective.

The plan was to estimate the ICC of the QIDS-SR ACASI score using mixed models regression methods based on week 0 data. With this estimate, a revised sample size computation based on 80% power would be completed. The power of our design also would be re-assessed, given the computed ICC and the current target sample size.

This pre-planned interim analysis took place in October, 2018. The report concluded that, "even having lost one site from the initial design, and having a second site with minimal enrollment, with the current target sample size of N=156, we would still have 80% power were we to accrue an average of 10 evaluable participants per each of the 12 remaining sites given the observed intracluster correlation and allowing for a 20% non-evaluability rate." Based on these results, the core study team made the decision not to alter the sample size target.

# 7 Statistical Considerations

## 7.1 Participant characteristics

Participant characteristics (gender, age, mode of transmission, severity of depression, CD4 and RNA levels and viral suppression status) at study entry will be summarized and compared between treatment arms using clustered measures appropriate for categorical (number, percent) and continuous (mean, standard deviation or median, $25^{th}$ and $75^{th}$ percentiles) outcomes. We will also note whether sites allowed parental waivers of consents.

## 7.2 Primary Analysis (Primary study objectives)

The primary analyses of the primary study objectives are planned as cluster-level analyses, where the unit of analysis is the site. Although this method does not use all the information in the data, this method is recommended when there are fewer than 15-20 clusters per group because individual-level analyses do not appear to perform robustly with small numbers of clusters (1). The cluster-level analyses will involve a two-stage procedure. The first stage will calculate a summary measure for each site – the mean for continuous outcome measures (QIDS-SR, CD4 count) and the proportion for dichotomous outcomes (response to treatment, remission, undetectable plasma HIV RNA). The second stage will perform a two-

sample t-test (with two-sided alpha =0.05) on the site-specific summary measures. Note that the alpha level will not be adjusted for multiple comparisons because the Week 24 QIDS-SR score has been designated as the primary efficacy outcome and the response, remission, CD4, and RNA outcomes are considered secondary efficacy outcomes. However, the interpretation of the results will comment on whether a multiple comparisons adjustment would change the conclusions. If the summary measures have a skewed distribution, a log transformation will be performed before doing the t-test. The primary analysis will focus on the outcome measures at Week 24 that are listed in Section 4.3 because the protocol team believes that it is important that the intervention demonstrate a lasting effect while a shorter term effect may not be as clinically meaningful. For the measures of depression outcomes at week 24 only the ACASI interview data will be used in the primary analysis.

### 7.3 Supplementary Analyses (Primary study objectives)

The following supplementary and exploratory analyses of primary study objectives will be carried out as appropriate.

#### 7.3.1 Non-parametric analyses

Methods based on the t-test have been shown to be highly robust to departures from the underlying assumptions. However, with small numbers of clusters per arm, the cluster-level t-test will be less robust to non-normality of the underlying distribution of cluster summaries, and it will be difficult to make a reliable assessment of non-normality. A non-parametric test that does not rely on a normality assumption, but which may be less powerful than the t-test, such as the Wilcoxon rank sum test or a permutation test (based on all the permissible allocations that were generated for the site randomization) will be performed to supplement the primary analysis.

#### 7.3.2 Time-oriented analyses

Secondary analyses that incorporate information from earlier time points (study entry, weeks 1, 6, and 12) as well as week 24 will also be conducted to supplement the results of the primary analyses. If ACASI data are not available for these time points (e.g., at weeks 1, 6 and 12, participants in the COMB-R group only completed the QIDS-SR on paper), data from the paper forms will be acceptable for analysis. The study team developed an algorithm to define the participant's baseline QIDS-SR value which prioritized ACASI data at week 0. Refer to the Analysis Implementation Plan (AIP) for details. The time-oriented analyses will include cross-sectional summaries at each time point, with either a non-parametric cluster-level comparison at each time point or the non-parametric Wei-Johnson test for change over time between groups. As noted above, cluster-level analyses are robust to departures from assumptions, but do not use all the information in the data, since the outcomes are measured at the individual participant level.

### 7.3.3 Individual-level analyses

Secondary analyses of the primary outcome measures will be conducted using individual-level regression methods, including mixed-effects linear regression models for continuous outcomes and generalized estimating equations (GEE) models for dichotomous outcomes, to supplement the primary cluster-level analyses for week 24 outcomes. These regression models also will be used to evaluate the influence of the COMB-R intervention on QIDS-SR trajectories over time. Such analyses can use all available outcome data, even for those participants with data missing for some time points.

### 7.3.4 Selection bias

To assess the potential for selection bias, the numbers and characteristics (gender, age, mode of transmission, severity of depression, and viral suppression status) of participants screening but not enrolling in the study will also be compared between treatment groups, and the characteristics of participants enrolling vs. not enrolling will be compared overall and within treatment arms as well. These comparisons are limited by the fact that data were only collected for those participants who first consented to enroll. To assess potential selection bias after the implementation of LOA#1, we will compare characteristics of participants enrolled before and after implementation of LOA#1, overall and within treatment arm.

### 7.3.5 Loss to follow-up

High rates of loss to follow-up before Week 24, particularly if differential between randomized arms might affect the ability of the study to answer its primary objective, i.e. power may be reduced and the interpretation of any difference between randomized arms may be complicated by potential differences between participants who do versus do not remain in follow-up, particularly if this also differs between randomized arms. Allowance for a loss to follow-up rate of up to 10% prior to Week 24 was built into the initial sample size/power considerations for the study, and the protocol (Schema and Section 9.5.1) specifies that the sample size may be increased if the loss-to-follow-up rate exceeds 10%. At the time of the interim analysis, the loss-to-follow-up rate at Week 24 was between 11% and 15%, depending on definition of visit windows; As such, in the revised power computations, non-evaluability rates up to 20% were considered. The study team chose not to alter the target sample size of N=156.

If more than 10% of participants are lost to follow-up before Week 24, or if a larger proportion of participants are lost to follow up in one arm (especially if the loss is due to serious adverse events, or suicide attempts), several analyses will be undertaken to explore the potential effects of loss to follow-up on the conclusions of the study, The characteristics at study entry (gender, age, mode of transmission, severity of depression and viral load suppression status) for participants who discontinued from the study before Week 24 will be compared between treatment arms. Characteristics of participants discontinuing vs. not discontinuing before Week 24 will also be compared overall and within treatment arms. For these

comparisons, cluster-level methods will be used as described above. In addition, the reasons for losses to follow-up will be assessed and compared between treatment arms, with special attention paid to reasons related to new, post-baseline grade 3 or higher adverse events, psychological hospitalizations and suicide attempts. Descriptive comparisons by site will be performed for sites with sufficient numbers of participants. In addition, the probability and timing of loss to follow-up will be summarized and compared using Kaplan-Meier plots and the log-rank test.

#### 7.3.6 Missing data

Sensitivity analyses will be conducted to explore the potential impact of missing data on the conclusions of the primary efficacy analyses. Of primary concern are missing QIDS-SR scores at Week 24 for participants who had low or decreasing QIDS-SR scores at time points before Week 24. The pattern of missing data will be described. The sensitivity analyses will be done in two ways: (a) as an extreme, by imputing missing QIDS-SR scores at Week 24 so as to minimize the difference between randomized groups (e.g., if the primary efficacy analysis among participants with available Week 24 QIDS-SR score suggests that COMB-R is better than ESC, impute a high (worse) Week 24 QIDS-SR score for COMB-R participants who do not have a Week 24 QIDS-SR score, and impute a low (better) Week 24 QIDS-SR score for ESC participants who do not have a Week 24 QIDS-SR score; and (b) more plausibly, by imputing missing QIDS-SR scores at Week 24 by treatment arm as the mean of the Week 24 QIDS-SR scores for participants in that treatment arm who also had similar QIDS-SR scores at earlier time points. In instances where similar patterns of earlier scores do not occur, we will consider imputing the mean values at week 24 for those participants with similar baseline values. Alternatively, rather than imputing the mean value, we could draw a value at random from the distribution of values for participants with similar patterns of missing data (or similar baseline values). Use of multiple imputation procedures will also be considered. The interpretation will need to be more cautious if the results of these analyses suggest different conclusions.

### 7.4 Secondary Objectives

Except as noted below, secondary objectives will be analyzed after the week 48 visit data collection has been completed and data have been prepared for analysis. This will occur approximately six months after the primary (week 24) data analysis.

#### 7.4.1 Adherence

Adherence to both HIV and depression medication and depression counseling sessions will be assessed as continuous and categorical measures. For example, adherence to HIV medications will be dichotomized based on the data distribution of the number of days when doses were missed. Likert scales for self-reports of adherence frequency (how often did you take your medications the way you were supposed to?) and efficacy (how good a job did you do at taking your medications?) will be

dichotomized or analyzed as an ordinal measure, based on the observed data distribution. The proportions of expected counseling sessions attended will likewise be summarized and analyzed as a continuous or categorical measure based on the distribution observed. For each dichotomized adherence assessment, the proportions of participants with good versus poor adherence will be compared using the cluster-randomized methods described above. Similarly, if it is decided that the proportions of expected sessions can be analyzed as a continuous variable, methods noted above will be used to compare the COMB-R group to the ESC group at Week 24 and Week 48. A limitation of the depression medication adherence analyses is that adherence data will only be available for participants who are taking depression medications, so this will be a non-randomized comparison.

#### 7.4.2 Week 48 depression outcomes

Week 48 depression treatment outcomes will be analyzed similarly to the Week 24 outcomes in order to ascertain whether any reported effects are maintained. For depression outcomes, only ACASI data will be included in analyses for weeks 24, 36 and 48. At weeks 1, 6 and 12, paper-completed QIDS-SR data may be included in the analysis for those participants in the COMB-R group, as these participants would not have completed an ACASI questionnaire at those time points. The study team developed an algorithm to define the participant's baseline QIDS-SR value which prioritized week 0 ACASI data and is explained in detail in the AIP.

#### 7.4.3 Effect modification

Analyses to adjust for imbalance and potential confounding, and assess effect modification will be performed using cluster-level methods and also using individual level regression methods (because cluster-level methods do not use all the information in the data). Each approach is described in turn below. Preliminary analyses will be completed at the time of the week 24 data analysis while final analyses will be completed once week 48 data are in the database and prepared for analysis.

In the cluster-level analyses, adjustment for baseline covariates (e.g., the QIDS-SR score at entry) will be done by modifying the first stage of analysis and then completing the second stage. In the first stage, an individual-level regression of the individual outcome measures (e.g., QIDS-SR scores at Week 24) on the baseline covariates will be performed, ignoring the clustering of the data and then the residuals between observed and predicted values from this model will be used as the summary measures in the second stage.

Specifically, in the first stage of the cluster-level analyses, all variables of interest except for the study group (ESC or COMB-R) will be entered into the regression mode ignoring clustering and the summary statistic for each cluster will be the residual based on comparison of the summary measure calculated from the observed values of the outcome measure in that cluster and the summary outcome measure calculated from the predicted values of the outcome measures from the model in the absence of an

intervention effect. For example, to adjust for baseline QIDS-SR score, the site-specific residual for the QIDS-SR score at Week 24 will be the difference between the observed site-specific mean QIDS-SR score and the predicted site-specific mean QIDS-SR score (mean of the predicted QIDS-SR scores for that site from a linear regression model of Week 24 QIDS-SR score on baseline QIDS-SR score); the site-specific residual for the response proportion will be the difference between the observed site-specific response proportion and the predicted site-specific response proportion (predicted response proportion for that site from a logistic regression model). Then, in the second stage, these site-specific residuals will be compared using the two-sample t-test or other test as described above. Analyses of effect modification for cluster-level covariates and individual level covariates will be performed using methods described by Hayes and Moulton (56).

As noted above, cluster-level analyses are robust to departures from assumptions, but do not use all the information in the data, since the outcomes are measured at the individual participant level. To supplement the cluster-level analyses of the secondary study objectives, additional analyses will be conducted using individual-level regression methods, including mixed-effects linear regression models for continuous outcomes and generalized estimating equations (GEE) models for dichotomous outcomes. These individual level analyses will facilitate adjustment for baseline covariates, longitudinal analyses of the outcome measures over the 48-week study period, and assessment of effect modification by individual-level covariates. For example, linear mixed models will be used to evaluate the influence of the COMB-R intervention on QIDS-SR trajectories over the course of the study.

#### 7.4.4 Behavioral risk

If depression decreases over the course of treatment, we would expect the level of high risk sex and alcohol/drug behaviors to decline. We will assess whether COMB-R treatment for depression is associated with reduced alcohol/drug and sex risk behaviors using similar methods as described for the primary study objectives. Dichotomized measures for levels of high risk behaviors at weeks 24 and 48 will be compared between the COMB-R and ESC arms.

#### 7.4.5 Treatment fidelity

The implementation fidelity at COMB-R sites and the counseling strategies and medication patterns at ESC sites will be summarized using descriptive statistics. The use of medication, the number of therapy visits, and treatment acceptability, for both clinicians and participants will be described and compared for both COMB-R and ESC using clustered approaches as earlier described. CBT adherence checklist and ESC therapy checklist results as well as the COMB-R MM checklist results will also be summarized.

#### 7.4.6 Safety

The frequency and pattern of new, post-baseline grade 3 or higher adverse events, psychological hospitalizations and suicide attempts will be described for both treatment arms and the proportion of

participants with each of these outcomes will be compared between arms using methods described for the primary study outcomes. Presumed relationships of adverse events to COMB-R and ESC counseling will be summarized and compared. Preliminary analyses will be completed at the time of the week 24 data analysis while final analyses will be completed once week 48 data have been collected.

### 7.5 Exploratory study objectives

Similar methods to those described for the primary and secondary study objectives will be used to assess whether inflammatory markers decrease with the COMB-R intervention compared to ESC and whether the inflammatory markers could moderate the treatment effects on the primary outcome measures.

## 8 Report contents

The primary statistical report will contain the following sections. Further details are provided in the AIP.

- CONSORT Diagram
- Enrollment, including potential selection bias
- Study status (also reported in ClinicalTrials.gov)
- Visit completion and loss-to-follow-up
- Baseline characteristics (also reported in ClinicalTrials.gov)
- Treatment status (reported in ClinicalTrials.gov)
  Note: The “treatment” in IMPAACT 2002 consists of the counseling interventions.
- Primary analysis of primary outcome measures (reported in ClinicalTrials.gov)
- Supplemental/exploratory analysis of primary outcome measures
- Analysis of secondary outcome measures (reported in ClinicalTrials.gov)
- Adverse events (reported in ClinicalTrials.gov)

## 9 References

(1) Hayes RJ, Moulton LH. Cluster Randomised Trials: Taylor & Francis; 2009.

## Tables and Figures

**Table 1. Stages of Medication Management (COMB-R)**

| Stage | Treatment | Medication Options |
|---|---|---|
| Stage 0 | No medication | N/A |
| Stage 1 | Monotherapy with SSRI | Fluoxetine, citalopram, sertraline, escitalopram, paroxetine<br><br>Partial responders after six weeks may receive augmentation with lithium, bupropion, or an atypical antipsychotic |
| Stage 2 | Monotherapy with 2nd SSRI | Fluoxetine, citalopram, sertraline, escitalopram, paroxetine<br><br>Partial responders after six weeks may receive augmentation with lithium, bupropion, or an atypical antipsychotic |
| Stage 3 | Monotherapy with non-SSRI | Venlafaxine, bupropion, mirtazapine, duloxetine<br><br>Partial responders after six weeks may receive augmentation with lithium, bupropion, or an atypical antipsychotic |
| Stage 4 | Combination treatment | Two antidepressants or antidepressant plus lithium |

**Table 2. CBT Components**

| | |
|---|---|
| a | Psychoeducation |
| b | Motivational interviewing strategies |
| c | Adherence training |
| d | Behavioral coping |
| e | Cognitive restructure and automatic thoughts |
| f | Problem solving |
| g | Wellness |
| h | Practice and application |
| i | Homework assignment |
| j | Relapse & Wellness plan |
| k | Safety plan |
| l | Booster sessions |
| m | Family communication |
| n | Contingency management |
| o | Emotional regulation |

**Table 3. Enhanced Standard of Care Approaches**

| | |
|---|---|
| 12 | Supportive/coping with stress |
| 12 | Focused problem solving |
| 13 | Cognitive behavioral |
| 14 | Interpersonal |
| 15 | Acceptance and commitment |
| 16 | Dialectal behavioral |
| 17 | Expressive or emotion focused |
| 18 | Motivational interviewing or motivational enhancement |
| 19 | Relaxation and/or mindfulness |
| 20 | Eclectic or “personalized” treatment |
| 21 | Substance use focused treatment |
| 99 | Other |