Supplemental Methods

The NP battery (Table 1) was designed in accordance with Frascati guidelines [1] to be sensitive to the cognitive domains most often affected in HIV disease. The tests were scored using published normative data that corrected for age, sex, education, and ethnicity, as appropriate. Moreover, the battery was relatively brief and repeatable, facilitating its use in a multi-center clinical trial that requires multiple assessments over time. Unique alternate forms of the learning/recall tests were available and counterbalanced versions were administered at baseline and 16 weeks.

Medical History and Examination

Clinicians used a standardized form to obtain a structured medical history of prior conditions including serious injuries, surgeries, and hospitalizations on a standardized form and to record medical and neurological symptoms, including their duration and intensity. Clinicians performed a comprehensive neurological examination at screening and Weeks 16 and 32 that included assessment of mental status, cranial nerves, motor, sensory and cerebellar function, reflexes and gait.

Neurocognitive Diagnosis

All participants were classified according to the Frascati diagnostic criteria for HAND [1], which includes HIV-associated dementia (HAD), mild neurocognitive disorder (MND) and asymptomatic neurocognitive impairment

(ANI). ANI represents a condition in which NP testing documents objective impairment relative to appropriate norms, but self-reported daily functioning is unaffected. These criteria also classify comorbid neurocognitive conditions such as minor head injury and substance abuse into three categories: incidental/mild, contributing/moderate and confounding/severe.

Psychiatric and substance abuse diagnoses were rendered using the Composite International Diagnostic Interview [CIDI; [2]] for DSM-IV (Mood, Psychotic, and Substance Use Disorder Modules). Severity of current (past week) depressive symptoms was ascertained via the Beck Depression Inventory II [3].

Adaptive Randomization

The principal goal of the adaptive randomization was to achieve balance between the treatment groups with respect to characteristics that might influence the primary outcome measure [4]. These characteristics were past ARV treatment history (naïve versus experienced), current CD4+ T-cell count (<200 or ≥200 cells/mm^3), hepatitis C virus serostatus, severity of NP impairment, and study site. The contributions of each of these factors towards balancing the groups were weighted differentially: severity of NP impairment was given a weight of three, study site a weight of two and past ARV treatment history, CD4 count and hepatitis C virus serostatus were given a weight of one each. For the first subject randomized, the probability of assignment to the CNS-T or comparison arm was 0.50. For subsequent participants, probability of assignment to the CNS-T or comparison arm depended on the balance between these

characteristics for previously randomized participants and was 0.90 for the arm which would make the groups more balanced and 0.10 for the arm which would make the groups less balanced according to the stratification variables and weighting listed above [4].

Resistance Genotyping and Phenotyping

ARV resistance testing was performed by Monogram Biosciences, Inc. (San Francisco) using the PhenoSense GT® assay (combination HIV phenotype and genotype) for participants who failed a previous ARV regimen, and by GenoSense® for participants who were naïve or had never failed a previous regimen [5]. A phenotypic sensitivity score (PSS) was computed for each subject for whom phenotypic testing was available. For each drug in the regimen, a sensitivity value between 0 and 1 was assigned, according to the rules and cut-offs in current use [5]. In addition, a relative PSS was calculated for each subject as the PSS averaged on the number of drugs in the regimen (not including ritonavir).

Study Treatments

All medications were FDA approved. Individualized ART regimen selection and randomization to the CNS-T or comparison arm were reconciled by using three procedures. First, an APC comprising an HIV treatment and resistance specialist, an ARV pharmacist and pharmacologist and an HIV neurologist reviewed each participant's treatment history, ARV resistance and comorbidities to establish a list regimen options designed to provide full virologic suppression for that participant. Each participant's PCP reviewed the regimen list to ensure

that all options were acceptable. Second, CPE ranking criteria were applied as published in 2007 [6]. To ensure that all participants were eligible to be randomized to either treatment arm and that CPE scores between the two arms did not overlap, each participant's regimen list was required to include both regimens with CPE ranks ≥ 2.0 and regimens with ranks ≤ 1. Third, key clinical factors (and site) were balanced across the two arms by using an adaptive computer randomization. PCPs then prescribed the study ARVs and participants visited the local study site to take their first doses under direct observation. Adherence data were collected throughout the study.

Adherence and Concomitant Medications

ARV adherence was assessed by self-report and determination of plasma ARV concentrations. The AIDS Clinical Trials Group (ACTG) Adherence to Anti-HIV Medications form [7] was applied to collect detailed information about adherence each day over the previous four days, and to estimate the number of doses that may have been missed in the prior two weeks. This measure also recorded the reasons for missed doses and asked about HIV-related symptom burden. Detailed information about past and present ARV exposure was captured on standardized case report forms at all study visits. Current ARV information included start and stop dates, dosage, date and time of last dose and whether the last dose was taken with food. All concomitant medications were recorded.

Interim Pharmacokinetic (PK) Evaluation

As part of the CNS-T strategy, each subject underwent an interim, sparsely-sampled plasma PK evaluation at Week 2. Participants in the comparison arm also underwent PK evaluations, but results and recommendations (see below) were not provided to the treating providers. For each subject, concentrations of any non-nucleoside reverse transcriptase inhibitors (NNRTI) or protease inhibitors (PI) in the regimen were measured. Nucleoside reverse transcriptase (NRTI) levels were not measured. The sparsely-sampled PK consisted of three measured concentrations: immediately before, and 2 and 4 hours after a clinically observed dose of medication. Samples were processed and shipped to a central laboratory.

Plasma ARV Concentration Targeting (PACT)

A Bayesian nonlinear curve fitting approach was used to estimate individual PK parameters from dosing and concentration information. These calculations used population PK parameters (and their variance) from previous PK studies. The calculated pharmacologic metrics (minimum concentration, maximum concentration and area-under-the-time-concentration-curve) were compared to population values and the percent of predicted values was calculated. The precision and bias of individual PK fits, as well as the deviation from typical population parameters, was quantified as the percent mean absolute error and percent mean error.

Interpretation of PK data was based on a comparison of the patient's achieved concentration to the population-predicted concentration (e.g., the 50th

percentile is the concentration at which 50% of the patients are above or below). ARV concentrations below a nominal threshold (25th percentile of predicted based on population PK) were considered suboptimal. Suboptimal drug concentrations were then considered in the context of self-reported medication adherence, other factors that might influence drug disposition, such as concomitant medications, and the concurrent virologic response. If the 2-week virologic response was not adequate (at least -1.0 log10 plasma viral load compared to baseline), the APC formulated recommendations ranging from adherence support to dose adjustment to consideration of alternative concomitant medications [8]. Alternatively, if the concentration was above a threshold (i.e., the 80th percentile), drug related toxicities and dose reduction were considered.

Study Monitoring

An independent data safety monitoring board (DSMB) evaluated the safety of the trial and the effectiveness of reporting annually. DSMB members were experts in ARV therapy, HIV neurology, neuropsychology and statistical analysis not involved in the study. The study was designated as low risk, since 1) all ARV agents used were FDA-approved or available through FDA-monitored expanded access programs and 2) PCPs limited the pool of regimens available for randomization. The DSMB reviewed aggregate data on recruitment, retention, demographics, data completeness and quality and blinded efficacy results for the two treatment arms, identified as arm “X" and arm “Y". Additional blinded data assessed between the two arms included demographics, differences in CPE

ranks, rates of plasma HIV suppression, incident neurocognitive dysfunction and rates of and reasons for changes in ARV drugs.

Supplemental Results

Virologic Suppression

Among 35 subjects with detectable plasma viral load at baseline, 19 (54.3%) achieved suppression by week 16. By treatment arm, 36.8% (7/19) in the CNS-T arm achieved suppression, while in the non-CNS-T arm, 75% (12/16) achieved suppression ($p = 0.06$). Regardless of treatment assignment, the improvement in adjusted GDS for study participants who achieved suppression (N=19; mean = -0.13; 95% CI -0.34, 0.07) did not differ significantly from that observed in participants who remained unsuppressed through week 16 (N=16; mean = -0.14; 95% CI -0.47, 0.20), corresponding to a small effect size (-0.06; 95% CI -0.72, 0.61; $p = 0.87$). There was no interaction between suppression status at week 16 and treatment arm ($p = 0.43$).

Resistance Testing

Among 13 participants (26.5%) with plasma viral loads <500 copies/mL prior to randomization, resistance testing could not be performed. For one additional participant with a plasma viral load of 1303 copies/mL, the resistance assay failed. Nine ARV-naïve participants underwent genotypic resistance testing only; one additional subject with minimal ARV exposure had genotypic testing, and for 2 previously performed genotyping was used. The remaining 23 participants had combined geno- and phenotypic testing done. For those with

phenotypic testing, the mean (SD) PSS was 3.4 (0.7) and the relative PSS was 0.98 (0.1), indicating that most participants were randomized to regimens with at least 3 susceptible drugs. The mean relative PSS scores did not differ between the CNS-T and comparison arms (1 vs 0.95, p = 0.19).

As Treated Analysis

At the Week 16 primary outcome visit, 42 of 49 participants (86%) remained on the regimen to which they had been randomized. Among those who changed, 2 stopped ART entirely, and 5 changed regimens. Of those who stopped, one did so because of side effects, and one relapsed with substance abuse. For those who changed regimens, one did so at the PCP's request, 3 because of side effects, and one because of virologic failure. Prior to randomization, the participant took darunavir, ritonavir, tenofovir and emtricitabine. He was randomized to the non-CNS-T arm and received saquinavir, ritonavir, tenofovir and lamivudine. Virologic failure (escape) was detected initially in CSF at week 6, and subsequently in plasma at week 10. With a subsequent reversion to the previous regimen and the addition of zidovudine, he achieved full viral suppression in both plasma and CSF by week 20.

Sensitivity Analyses

A sensitivity analysis was performed to evaluate potential biases introduced by participants who dropped out of the study prior to week 16 (N=10). Missing primary outcome (GDS) values for these participants were replaced with the worst outcome (GDS = 5). The 10 imputed observations were added to the data and the primary analysis was repeated. Assumptions of normality in $\log_{10}$ of

GDS change and homogeneity of variances were satisfied by Shapiro-Wilk and Levene tests. The analysis yielded a point estimate of 0.16 for the difference in average GDS change on the log scale with 95% CI = (-0.09, 0.42). The corresponding effect size between the two arms, measured as standardized mean difference, was small, ES = 0.33, with 95% CI = (-0.19, 0.84). The difference between arms in change in $\log_{10}$ GDS between baseline and week 16 was not significant (p = 0.21, adjusted for baseline GDS; p = 0.20, unadjusted). An additional sensitivity analysis assumed no change (GDS change = 0) in these 10 participants. Since the assumption of normality was not met for non-transformed or log-transformed values, non-parametric methods were used and yielded no significant treatment effect (p = 0.81, adjusted for baseline GDS; p = 0.26, unadjusted). Additional analyses using Cook's distance measures revealed no potential outliers in the 49 observations.

Sample Size Calculation for Future Trials

Additional power analyses were conducted to estimate the sample size needed to detect a population effect size equal to the observed effect size in this study. The analysis used two observed effect size estimates, the first based on the preplanned, intent-to-treat outcome of unadjusted GDS change from baseline to week 16 (d = 0.09), and the second on the practice effect-adjusted GDS change as previously described (d = 0.25). The sample sizes required to obtain a conditional power of 80% were 3876 and 504, respectively, compared to the planned sample size of 120, and the actual sample size of 59. These power

estimates should be considered in the context of the confidence intervals for the observed effect sizes.

Impact of on-study regimen changes on CPE

For two participants who changed from the randomized regimen, the CPE scores of the replacement regimens corresponded to a switch in treatment arm: one from the CNS-T arm to the comparison arm, and one from the comparison arm to the CNS-T arm. The CPE scores of the remaining three replacement regimens were similar or identical to their randomized regimens (1 CNS-T, 2 non-CNS-T). Considering the 42 participants who remained on the regimen to which they had been randomized, the arms did not differ significantly in the mean (SD) GDS change, adjusted for baseline GDS: -0.24 (0.54) vs -0.17 (0.41); treatment effect size -0.03, (95% CI -0.64, 0.58).

Plasma Pharmacokinetics

All but one of the 49 participants in the intent-to-treat analysis completed the sparsely-sampled plasma PK evaluation at Week 2 as described above. Following guidelines described above (see "Plasma ARV Concentration Targeting"), the APC reviewed ARV plasma drug concentrations and formulated recommendations ranging from adherence support to dose adjustment to replacing concomitant interacting medications. The proportions of study participants with suboptimal ARV plasma concentrations as described above were similar for the two study arms: CNS-T, 6/25 (24%) vs. non-CNS-T, 6/23 (26%).

Other Secondary Analyses

Separate regression analyses were performed to evaluate potential treatment effects on individual NP domains, including learning, recall, verbal, executive function, speed of information processing, working memory, motor. Only the learning domain approached statistical significance ($p = 0.08$), with greater improvement in the CNS-T arm than the non-CNS-T arm.

Severe neurocognitive confounds that made participants ineligible for a diagnosis of HAND [1] were discovered after randomization in four participants (7%), 3 in the CNS-T and 1 in the non-CNS-T group. Excluding these participants from the analysis did not substantially alter the pattern of findings. An additional 13 participants (CNS-T, 8; non-CNS-T, 5) had moderate/contributing NP confounds. Overall, the distribution of confounding conditions did not differ in the CNS-T vs non-CNS-T groups ($p=0.55$).

Published data on the CNS penetration of ARVs continued to accumulate after this trial began, and an updated version of the CPE ranking system was disseminated in 2010 [9]. CPE rankings according to the new system still differed substantially between arms (10 [9-13] vs. 7 [5-9]) but did not correlate with practice effect-adjusted GDS change ($r = 0.043$; $p = 0.77$).

Among 48 participants in the as-treated sample who had plasma pharmacokinetic testing at study week 2 (one subject missed the week 2 visit), 29 (60%) had drug concentrations within the predicted range (25th-75th percentile). In the 12 participants (25%) who had levels below the 25th percentile, counseling to improve adherence was given in 8, and dose increases

were recommended and undertaken for the remaining 4. In the 7 participants who had concentrations above the 75th percentile, 7 showed no evidence of toxicity, and no dose change recommendations were made. One subject had significant diarrhea that improved after a reduction in ritonavir dose was recommended.

Supplemental References

1. Antinori A, Arendt G, Becker JT, Brew BJ, Byrd DA, Cherner M*, et al.* Updated research nosology for HIV-associated neurocognitive disorders. *Neurology* 2007,**69**:1789-1799.

2. WHO. Composite International Diagnostic Interview, Version 2.1. In. Geneva: World Health Organization; 1997.

3. Beck AT, Steer RA, Brown GK. Beck Depression Inventory In. Second ed. San Antonio, TX: The Psychological Corporation; 1996.

4. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 1975,**31**:103-115.

5. Swanstrom R, Bosch RJ, Katzenstein D, Cheng H, Jiang H, Hellmann N*, et al.* Weighted phenotypic susceptibility scores are predictive of the HIV-1 RNA response in protease inhibitor-experienced HIV-1-infected subjects. *J Infect Dis* 2004,**190**:886-893.

6. Letendre S, Ellis RJ, Best BB, Bhatt A, Marquie-Beck J, LeBlanc S*, et al.* Penetration and Effectiveness of Antiretroviral Therapy in the Central Nervous System. *Anti-Inflammatory & Anti-Allergy Agents in Medicinal Chemistry* 2009, **8**:169-183.

7. Chesney MA, Ickovics JR, Chambers DB, Gifford AL, Neidig J, Zwickl B*, et al.* Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: the AACTG adherence instruments. Patient Care Committee & Adherence Working Group of the Outcomes Committee of the Adult AIDS Clinical Trials Group (AACTG). *AIDS Care* 2000,**12**:255-266.

8. Best BM, Goicoechea M, Witt MD, Miller L, Daar ES, Diamond C*, et al.* A randomized controlled trial of therapeutic drug monitoring in treatment-naive and -experienced HIV-1-infected patients. *J Acquir Immune Defic Syndr* 2007,**46**:433-442.

9. Letendre SL, Ellis RJ, Ances BM, McCutchan JA. Neurologic complications of HIV disease and their treatment. *Top HIV Med* 2010,**18**:45-55.

Supplemental Table 1. Full NP Test Battery

Premorbid Estimate of Intellectual Functioning

Wide Range Achievement Test - 3rd Edition

Reading subtest (baseline only)

Speed of Information Processing

WAIS-III Digit Symbol and Symbol Search

Trail Making Test, Part A

Color-Word Test (Color Naming) – D-KEFS Version

Learning and Delayed Recall (2 domains)

Hopkins Verbal Learning Test-Revised

Brief Visuospatial Memory Test-Revised

Abstraction/Executive Functioning

Wisconsin Card Sorting Test (64-item)

Trail Making Test, Part B

Color-Word Test (Interference Trial) – D-KEFS Version

Verbal Fluency

Controlled Oral Word Association Test

Category Fluency (Animals), Action Fluency

Attention/Working Memory

WMS-III – Spatial Span

PASAT (50-item)

Motor

Grooved Pegboard Test