**Supplemental Appendix**

**to**

**“No Neurocognitive Advantage for Immediate Antiretroviral Treatment in adults with greater than 500 CD4 T Cell Counts”**

Edwina J Wright, Birgit Grund, Kevin Robertson, Lucette Cysique, B J Brew, Gary Collins, Mollie Poehlman-Roediger, Michael J Vjecha, Augusto Penalva de Oliveira, Barbara Standridge, Cate Carey, Anchalee Avhingsanon, Per O Jansson, Eric Florence, Jens Lundgren, Alejandro Arenas-Pinto, Nicolas J Mueller, Alan Winston, Moses Supercharger Nsubuga, Luxshimi Lal, Richard W. Price for the INSIGHT START Neurology Substudy Group

**Table of Contents**

Statistical Methods (Expanded) 2

Table S1. Neuropsychological test scores at baseline 5

Table S2A. Changes in the QNPZ-8 from baseline 6

Table S2B. Changes in Grooved Pegboard z-scores 7

Table S2C. Changes in Finger Tapping z-scores 8

Table S2D. Changes in WAIS-III Digit Symbol z-scores 9

Table S2E. Changes in Semantic Verbal Fluency z-scores 10

Table S2F. Changes in Color Trails 1 z-scores 11

Table S2G. Changes in Color Trails 2 z-scores 12

Table S2H. Changes in HVLT-R Learning z-scores 13

Table S2I. Changes in HVLT-R Delayed Recall z-scores 14

Table S3. Prevalence of neurocognitive impairment (NCI) 15

Table S4. Changes in the Global Deficit Score (GDS) 16

Table S5. Changes in the Center for Epidemiological Studies Depression scale (CES-D) 17

Table S6. Prevalence of depression (CES-D > 16) 18

Table S7. Baseline characteristics by pre-specified ART regimen (EFV versus no EFV) 19

***Statistical Methods (Expanded)***

The primary outcome was “change in the QNPZ-8 score”, where the QNPZ-8 score is the average over the z-scores for the 8 neurocognitive tests in the battery: Grooved Peg Board, Finger Tapping, WAIS-III Digit Symbol, Semantic Verbal Fluency, Color Trails 1, Color Trails 2, HVLT-R Learning, HVLT-R Delayed Recall. The Grooved Peg Board and Finger Tapping tests were repeated with the dominant and non-dominant hands; the z-scores for these two tests were calculated by averaging the z-scores from the dominant and non-dominant hands. The change in the QNPZ-8 score was calculated as follows: for each individual test, and each follow-up visit, the change in z-scores was calculated as difference between the follow-up visit z-score minus the baseline z-score; “change in QNPZ-8” from baseline to a given follow-up visit was then calculated as average of the “change in z-scores” for the 8 individual tests.

The primary analysis for the Neurology substudy was an intent-to-treat comparison between the immediate and deferred ART groups for changes in QNPZ-8 from baseline through follow-up, using a longitudinal mixed model with random participant-specific intercept. The response in the mixed model was “change in QNPZ-8 from baseline”, fixed effects included an indicator variable for treatment group, visit number (as categorical variable), and baseline QNPZ-8 scores (continuous variable).

The sample size of 600 participants was estimated to detect an average treatment difference in the change in QNPZ-8 scores between the two study arms of 0.13 with 80% power at a 5% significance level, using the primary analysis method described above.

For all analyses, follow-up was censored at each participant’s last visit prior to May 27, 2015, when the parent START study was unblinded. To illustrate the differential use of ART in the immediate and deferred groups and its effect on CD4 cell counts and HIV RNA levels through follow-up, the proportion of participants using ART, the proportion of participants with HIV RNA < 200 copies/mL and the mean change in CD4 cell counts were summarized by treatment group in 4-month intervals. The treatment difference in change in CD4 cell counts through follow-up was estimated in a longitudinal mixed model adjusted for visit and baseline CD4.

In addition to the primary analysis which estimated the treatment effect as a weighted average over all follow-up visits, we also compared treatment groups for changes in QNPZ-8 from baseline through month 12 only, using similar longitudinal mixed models. By design, participants in the immediate group were to initiate ART at randomization, while few participants in the deferred group initiated ART within the first year. Therefore, the difference between the immediate and deferred ART groups over the first 12 months is a better approximation of the effect of ART versus no ART than the primary comparison (which includes the full follow-up), as only few participants in the deferred ART group initiated ART during the first year (ART use is illustrated in Figure 2A). We performed similar intent-to-treat comparisons of changes in z-scores for each of the 8 tests as planned per protocol. Within each treatment group, changes in QNPZ-8 and individual z-scores from baseline to annual visits were summarized by means with 95% confidence intervals (CIs). Treatment groups were compared for changes from baseline to each visit using t-tests in linear regression models adjusted for baseline scores; specifically, these linear regression models were Analysis of Covariance (ANCOVA) models that included “change from baseline to visit” as response, the treatment group indicator as treatment factor, and the baseline score as covariate. We compared treatment groups for changes in the prevalence of neurocognitive impairment (NCI) from baseline through follow-up using generalized estimating equations (GEE) for binary responses with a logit link function as in logistic regression; these GEE models included the treatment group indicator, an indicator for follow-up visits (taking the value 0 for the baseline visit, 1 for follow-up visits), their interaction effect, and visit number (categorical variable). In this model, the interaction between the treatment group indicator and the indicator for follow-up visits estimates the log of the treatment effect, and the treatment effect is modeled as a ratio (immediate/deferred) of odds ratios (odds of NCI during follow-up/odds of NCI at baseline). The p-value for the interaction effect thus tests whether the prevalence of NCI through follow-up is equal between treatment groups, after adjustment for a possible difference in prevalence at baseline, taking into account the within-subject correlation across visits. We used similar GEE models to compare the immediate versus the deferred group for changes from baseline through follow-up in the prevalence of depression (CES-D > 16). In order to compare the prevalence of NCI (or depression) between treatment groups at each visit, we used Chi-squared tests for proportions. Because the prevalence of NCI differed between treatment groups at baseline, we also compared the prevalence of NCI between treatment groups at each visit after adjustment for differences in baseline prevalence, by modeling the prevalence of NCI at baseline and at the given follow-up visit using a GEE model for binary responses that included the treatment group indicator, and indicator for visit (baseline versus the given follow-up visit), and their interaction. As described above, the p-value for the interaction effect compares the “change in prevalence of NCI from baseline” between treatment groups, and thus takes into account differences in the prevalence at baseline when evaluating the treatment effect on the prevalence of NCI at the given follow-up visit.

We compared treatment groups for changes in GDS using longitudinal mixed models, which included “change in GDS from baseline” as response, a subject-specific intercept as random effect, and the following fixed effects: the treatment indicator, visit number (as categorical variable) and baseline GDS (continuous).

To assess the effect of ART versus strictly untreated HIV, we also compared the immediate group (excluding participants who did not start ART within the first year) versus the deferred groups censored at ART start for changes in QNPZ-8 and individual z-scores; this comparison is not protected by randomization.

Subgroup analyses for the primary endpoints were performed to determine whether the treatment effect differed across baseline characteristics. The homogeneity of the treatment effect across subgroups was assessed by testing for interaction between the subgroup variable and treatment group indicator in longitudinal mixed models; when possible, the continuous subgroup variable (e.g., age or log10 HIV RNA) was used to test for homogeneity. We investigated 24 subgroups, listed in the footnote to figure 3. We provided the p-values for the interaction effects without adjustment for multiple comparisons. In order to adjust for multiple comparisons, we applied the Benjamini-Hochberg FDR method to the 24 interaction p-values for the investigated subgroups.

We compared treatment groups for changes in CES-D scores using longitudinal mixed models with participant-specific random intercept, and the following fixed effects: treatment group indicator, visit (categorical variable), and baseline CES-D.

Analyses were performed with SAS version 9.4 (SAS Institute, Cary, North Carolina, United States) and R version 3.22 All p-values are two-sided; p<0.05 denotes significance.

**Table S1.** Neuropsychological test scores at baseline.

|  |  |  |
| --- | --- | --- |
|  |  | **Mean (SD)** |
| **Test** |   | **Immediate Group(n= 291)**  | **Deferred Group(n= 301)** | **Total(n= 592)** |
| Grooved Peg Board (seconds) |   |   |   |   |
|     Dominant hand |   | 67.7 (12.0) | 69.2 (15.4) | 68.3 (13.8) |
|     Non-dominant hand |   | 74.4 (16.2) | 76.5 (22.1) | 75.2 (19.5) |
|     Average |   | 71.1 (13.1) | 72.8 (18.1) | 72.0 (15.9) |
| Finger Tapping Test (no. of taps) |   |   |   |   |
|     Dominant hand |   | 47.2 (10.0) | 46.0 (9.6) | 46.6 (9.8) |
|     Non-dominant hand |   | 43.2 (8.7) | 42.4 (8.4) | 42.8 (8.5) |
|     Average |   | 45.2 (8.8) | 44.2 (8.5) | 44.7 (8.7) |
| WAIS-III Digit Symbol (no. correct) |   | 65.8 (16.6) | 66.2 (17.1) | 66.0 (16.8) |
| Semantic Verbal Fluency (no. correct) |   | 15.0 (4.3) | 14.3 (4.3) | 14.7 (4.3) |
| Color Trails 1 (seconds) |   | 41.3 (19.9) | 40.1 (16.2) | 40.7 (18.1) |
| Color Trails 2 (seconds) |   | 78.0 (30.1) | 78.4 (27.6) | 78.2 (28.8) |
| HVLT-R Learning Trials (no. correct) |   | 26.3 (4.0) | 26.1 (4.4) | 26.2 (4.2) |
| HVLT-R Delayed Recall (no. correct) |   |  9.5 (1.9) |  9.3 (2.1) |  9.4 (2.0) |
|  |  |  |  |  |
| QNPZ-8† |   | 0.03 (0.57) | -0.03 (0.61) | 0.00 (0.59) |
| † QNPZ-8 = quantitative neuropsychological performance z-score, the average of the z-scores for the eight tests for each participant. Z-scores for each test were calculated by subtracting the mean and dividing by the SD of all study participants at baseline; for the finger tapping and grooved pegboard tests, z-scores for the dominant and non-dominant hands were calculated separately and averaged before calculating the QNPZ-8. Six cognitive domains are covered with this test battery: speed of information processing (Color Trails 1 and WAIS III Digit Symbol); mental flexibility (Color Trails 2); verbal learning and memory (HTLV\_R Total Learning and Delayed); verbal fluency (Semantic Fluency); fine motor skills and motor speed (dominant and non-dominant hand Grooved Pegboard, and dominant and non-dominant hand Finger Tapping). |

**Table S2A.** Changes in the QNPZ-8 summary score, from baseline through follow-up.

|  |
| --- |
|  |
| **QNPZ-8, Change from Baseline** |
|   | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred)†** |
|   |  |  |  |
| **Test and Visit** | **N** | **Mean** | **N** | **Mean** | **Difference** | **95% CI** | **P-value** |
| **QNPZ-8 score** |  |  |  |  |  |  |  |
| Month 4 | 273 | 0.115 | 282 | 0.125 | -0.010 | (-0.065, 0.045) | 0.72 |
| Month 8 | 269 | 0.137 | 272 | 0.162 | -0.026 | (-0.081, 0.030) | 0.36 |
| Month 12 | 273 | 0.215 | 273 | 0.235 | -0.019 | (-0.076, 0.037) | 0.50 |
| Month 24 | 265 | 0.222 | 259 | 0.198 | 0.024 | (-0.048, 0.096) | 0.51 |
| Month 36 | 247 | 0.252 | 252 | 0.253 | -0.002 | (-0.065, 0.062) | 0.96 |
| Month 48 | 151 | 0.219 | 136 | 0.226 | -0.007 | (-0.105, 0.091) | 0.89 |
| Month 60 | 30 | 0.306 | 34 | 0.128 | 0.178 | (-0.016, 0.372) | 0.08 |
|  |   |   |   |   |   |   |   |
| Overall‡ |   | 0.194 |   | 0.212 | -0.018 | (-0.062, 0.027) | 0.44 |
| Through Month 12 |   | 0.154 |   | 0.173 | -0.019 | (-0.064, 0.026) | 0.41 |

**†**At each visit, the means and the treatment difference were estimated in a linear regression model that included the treatment group indicator and the baseline QNPZ-8 score. The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline QNPZ-8 score, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

‡ Primary analysis. The treatment difference was estimated in a longitudinal mixed model, and represents a weighted average of the treatment differences at each visit.

**Table S2B.** Changes in the Grooved Pegboard z-score, from baseline through follow-up.

|  |
| --- |
|  |
| **Grooved Pegboard z-score, Change from Baseline** |
|   | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred) †** |
|   |  |  |  |
| **Visit** | N | Mean | N | Mean | Difference | 95% CI | P-value |
| Month 4 | 273 | 0.140 | 281 | 0.110 | 0.029 | (-0.056, 0.115) | 0.50 |
| Month 8 | 269 | 0.228 | 272 | 0.195 | 0.033 | (-0.048, 0.115) | 0.42 |
| Month 12 | 273 | 0.271 | 273 | 0.265 | 0.007 | (-0.081, 0.094) | 0.88 |
| Month 24 | 265 | 0.304 | 259 | 0.227 | 0.077 | (-0.060, 0.213) | 0.27 |
| Month 36 | 247 | 0.357 | 252 | 0.296 | 0.061 | (-0.028, 0.150) | 0.18 |
| Month 48 | 151 | 0.282 | 136 | 0.185 | 0.097 | (-0.120, 0.313) | 0.38 |
| Month 60 | 30 | 0.271 | 34 | 0.218 | 0.053 | (-0.195, 0.301) | 0.68 |
|  |   |   |   |   |   |   |   |
| Overall |   | 0.263 |   | 0.243 | 0.021 | (-0.041, 0.082) | 0.52 |
| Through Month 12 |   | 0.214 |   | 0.190 | 0.023 | (-0.041, 0.087) | 0.47 |

**†**At each visit, the means and the treatment difference were estimated in a linear regression model that included the treatment group indicator and the baseline z-score. The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline z-score, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S2C.** Changes in the Finger Tapping z-score, from baseline through follow-up.

|  |
| --- |
|  |
| **Finger Tapping z-score, Change from Baseline** |
|   | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred) †** |
|   |  |  |  |
| **Visit** | N | Mean | N | Mean | Difference | 95% CI | P-value |
| Month 4 | 272 | -0.026 | 281 | 0.036 | -0.062 | (-0.171, 0.046) | 0.26 |
| Month 8 | 268 | 0.055 | 272 | 0.023 | 0.032 | (-0.082, 0.147) | 0.58 |
| Month 12 | 271 | 0.060 | 272 | 0.019 | 0.041 | (-0.056, 0.139) | 0.41 |
| Month 24 | 263 | 0.144 | 259 | 0.063 | 0.081 | (-0.025, 0.187) | 0.13 |
| Month 36 | 247 | 0.200 | 252 | 0.102 | 0.098 | (-0.017, 0.212) | 0.10 |
| Month 48 | 151 | 0.061 | 136 | 0.060 | 0.000 | (-0.146, 0.147) | 1.00 |
| Month 60 | 30 | -0.035 | 34 | -0.153 | 0.118 | (-0.209, 0.444) | 0.48 |
|  |   |   |   |   |   |   |   |
| Overall |   | 0.064 |   | 0.057 | 0.007 | (-0.075, 0.088) | 0.88 |
| Through Month 12 |   | 0.032 |   | 0.028 | 0.004 | (-0.081, 0.089) | 0.92 |

**†**At each visit, the means and the treatment difference were estimated in a linear regression model that included the treatment group indicator and the baseline z-score. The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline z-score, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S2D.** Changes in the WAIS-III Digit Symbol test z-score, from baseline through follow-up.

|  |
| --- |
|  |
| **WAIS-III Digit Symbol Test z-score, Change from Baseline** |
|   | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred) †** |
|   |  |  |  |
| **Visit** | N | Mean | N | Mean | Difference | 95% C.I | P-value |
| Month 4 | 273 | 0.123 | 282 | 0.251 | -0.127 | (-0.234, -0.021) | 0.02 |
| Month 8 | 269 | 0.267 | 272 | 0.386 | -0.118 | (-0.233, -0.004) | 0.04 |
| Month 12 | 273 | 0.381 | 273 | 0.521 | -0.140 | (-0.261, -0.018) | 0.02 |
| Month 24 | 265 | 0.407 | 259 | 0.499 | -0.091 | (-0.223, 0.040) | 0.17 |
| Month 36 | 246 | 0.382 | 251 | 0.555 | -0.172 | (-0.306, -0.039) | 0.01 |
| Month 48 | 151 | 0.352 | 136 | 0.474 | -0.122 | (-0.326, 0.081) | 0.24 |
| Month 60 | 30 | 0.428 | 34 | 0.425 | 0.003 | (-0.436, 0.441) | 1.00 |
|  |   |   |   |   |   |   |   |
| Overall |   | 0.311 |   | 0.433 | -0.122 | (-0.208, -0.036) | 0.005 |
| Through Month 12 |   | 0.253 |   | 0.380 | -0.127 | (-0.217, -0.038) | 0.005 |

**†**At each visit, the means and the treatment difference were estimated in a linear regression model that included the treatment group indicator and the baseline z-score. The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline z-score, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S2E.** Changes in the Semantic Verbal Fluency z-score, from baseline through follow-up.

|  |
| --- |
|  |
| **Semantic Verbal Fluency z-score, Change from Baseline** |
|   | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred) †** |
|   |  |  |  |
| **Visit** | N | Mean | N | Mean | Difference | 95% C.I | P-value |
| Month 4 | 273 | -0.041 | 282 | 0.019 | -0.060 | (-0.209, 0.088) | 0.43 |
| Month 8 | 269 | -0.022 | 272 | -0.039 | 0.017 | (-0.144, 0.177) | 0.84 |
| Month 12 | 272 | -0.027 | 272 | -0.035 | 0.008 | (-0.150, 0.165) | 0.93 |
| Month 24 | 265 | -0.062 | 259 | 0.021 | -0.083 | (-0.249, 0.083) | 0.33 |
| Month 36 | 247 | -0.026 | 252 | -0.046 | 0.020 | (-0.138, 0.177) | 0.81 |
| Month 48 | 151 | -0.036 | 136 | -0.064 | 0.028 | (-0.176, 0.233) | 0.79 |
| Month 60 | 30 | 0.056 | 34 | 0.070 | -0.014 | (-0.454, 0.426) | 0.95 |
|  |   |   |   |   |   |   |   |
| Overall |   | -0.039 |   | -0.034 | -0.005 | (-0.105, 0.095) | 0.92 |
| Through Month 12 |   | -0.034 |   | -0.020 | -0.013 | (-0.126, 0.099) | 0.82 |

**†**At each visit, the means and the treatment difference were estimated in a linear regression model that included the treatment group indicator and the baseline z-score. The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline z-score, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S2F.** Changes in the Color Trails 1 z-score, from baseline through follow-up.

|  |
| --- |
|  |
| **Color Trails 1 z-score, Change from Baseline** |
|   | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred) †** |
|   |  |  |  |
| **Visit** | N | Mean | N | Mean | Difference | 95% C.I | P-value |
| Month 4 | 273 | 0.250 | 282 | 0.153 | 0.097 | (-0.012, 0.207) | 0.08 |
| Month 8 | 269 | 0.305 | 272 | 0.299 | 0.006 | (-0.097, 0.110) | 0.90 |
| Month 12 | 273 | 0.350 | 272 | 0.366 | -0.016 | (-0.109, 0.077) | 0.74 |
| Month 24 | 265 | 0.365 | 258 | 0.352 | 0.013 | (-0.073, 0.099) | 0.76 |
| Month 36 | 247 | 0.361 | 252 | 0.347 | 0.014 | (-0.082, 0.110) | 0.78 |
| Month 48 | 151 | 0.309 | 135 | 0.370 | -0.061 | (-0.194, 0.072) | 0.37 |
| Month 60 | 30 | 0.274 | 34 | 0.164 | 0.111 | (-0.160, 0.382) | 0.43 |
|  |   |   |   |   |   |   |   |
| Overall |   | 0.325 |   | 0.297 | 0.028 | (-0.046, 0.102) | 0.46 |
| Through Month 12 |   | 0.296 |   | 0.272 | 0.024 | (-0.059, 0.107) | 0.57 |

**†**At each visit, the means and the treatment difference were estimated in a linear regression model that included the treatment group indicator and the baseline z-score. The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline z-score, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S2G.** Changes in the Color Trails 2 z-score, from baseline through follow-up.

|  |
| --- |
|  |
| **Color Trails 2 z-score, Change from Baseline** |
|   | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred) †** |
|   |  |  |  |
| **Visit** | N | Mean | N | Mean | Difference | 95% C.I | P-value |
| Month 4 | 273 | 0.131 | 281 | 0.137 | -0.006 | (-0.109, 0.097) | 0.91 |
| Month 8 | 269 | 0.235 | 272 | 0.279 | -0.044 | (-0.138, 0.050) | 0.36 |
| Month 12 | 273 | 0.366 | 272 | 0.390 | -0.023 | (-0.116, 0.069) | 0.62 |
| Month 24 | 265 | 0.311 | 258 | 0.353 | -0.042 | (-0.134, 0.049) | 0.37 |
| Month 36 | 247 | 0.356 | 252 | 0.359 | -0.003 | (-0.094, 0.087) | 0.95 |
| Month 48 | 151 | 0.382 | 135 | 0.369 | 0.013 | (-0.100, 0.126) | 0.82 |
| Month 60 | 30 | 0.451 | 34 | 0.006 | 0.444 | (0.119, 0.770) | 0.010 |
|  |   |   |   |   |   |   |   |
| Overall |   | 0.287 |   | 0.290 | -0.003 | (-0.070, 0.065) | 0.94 |
| Through Month 12 |   | 0.238 |   | 0.267 | -0.030 | (-0.105, 0.046) | 0.44 |

**†**At each visit, the means and the treatment difference were estimated in a linear regression model that included the treatment group indicator and the baseline z-score. The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline z-score, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S2H.** Changes in the HVLT-R Learning Trial z-score, from baseline through follow-up.

|  |
| --- |
|  |
| **HVLT-R Learning Trial z-score, Change from Baseline** |
|   | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred) †** |
|   |  |  |  |
| **Visit** | N | Mean | N | Mean | Difference | 95% C.I | P-value |
| Month 4 | 273 | 0.249 | 282 | 0.155 | 0.094 | (-0.049, 0.238) | 0.20 |
| Month 8 | 269 | 0.133 | 272 | 0.070 | 0.063 | (-0.089, 0.215) | 0.42 |
| Month 12 | 273 | 0.249 | 273 | 0.177 | 0.072 | (-0.089, 0.233) | 0.38 |
| Month 24 | 265 | 0.272 | 259 | 0.108 | 0.165 | (-0.004, 0.334) | 0.06 |
| Month 36 | 246 | 0.374 | 251 | 0.265 | 0.109 | (-0.064, 0.281) | 0.22 |
| Month 48 | 151 | 0.324 | 136 | 0.395 | -0.072 | (-0.310, 0.167) | 0.56 |
| Month 60 | 30 | 0.573 | 34 | 0.307 | 0.266 | (-0.245, 0.778) | 0.31 |
|  |   |   |   |   |   |   |   |
| Overall |   | 0.312 |   | 0.211 | 0.100 | (-0.014, 0.214) | 0.08 |
| Through Month 12 |   | 0.209 |   | 0.126 | 0.083 | (-0.034, 0.201) | 0.17 |

**†**At each visit, the means and the treatment difference were estimated in a linear regression model that included the treatment group indicator and the baseline z-score. The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline z-score, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S2I.** Changes in the HVLT-R Delayed Recall z-score, from baseline through follow-up.

|  |
| --- |
|  |
| **HVLT-R Delayed Recall z-score, Change from Baseline** |
|   | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred) †** |
|   |  |  |  |
| **Visit** | N | Mean | N | Mean | Difference | 95% C.I | P-value |
| Month 4 | 273 | 0.151 | 282 | 0.087 | 0.063 | (-0.083, 0.210) | 0.40 |
| Month 8 | 269 | -0.030 | 271 | 0.006 | -0.036 | (-0.185, 0.113) | 0.64 |
| Month 12 | 273 | 0.119 | 273 | 0.126 | -0.007 | (-0.158, 0.144) | 0.93 |
| Month 24 | 265 | 0.077 | 259 | -0.062 | 0.140 | (-0.036, 0.315) | 0.12 |
| Month 36 | 246 | 0.073 | 251 | 0.090 | -0.017 | (-0.195, 0.161) | 0.85 |
| Month 48 | 151 | 0.058 | 136 | 0.100 | -0.042 | (-0.258, 0.173) | 0.70 |
| Month 60 | 29 | 0.361 | 34 | 0.042 | 0.319 | (-0.223, 0.862) | 0.25 |
|  |   |   |   |   |   |   |   |
| Overall |   | 0.084 |   | 0.067 | 0.017 | (-0.093, 0.126) | 0.77 |
| Through Month 12 |   | 0.077 |   | 0.064 | 0.013 | (-0.100, 0.127) | 0.82 |

**†**At each visit, the means and the treatment difference were estimated in a linear regression model that included the treatment group indicator and the baseline z-score. The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline z-score, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S3.** Number and percent of participants with mild neurocognitive impairment (NCI), defined as having z-scores < -1 in two or more of the 6 tested domains.

|  |  |
| --- | --- |
|  |  |
|  | **Immediate Group** | **Deferred Group** |  |  |
| **Visit** | **No. of pts in group** | **N (%) of ptswith NCI** | **No. of pts in group** | **N (%) of ptswith NCI** | **P-value†** | **P-value,adj.§** |
| Baseline | 291 | 50 (17.2) | 301 | 67 (22.3) |   |  |
| Month 4 | 273 | 37 (13.6) | 282 | 56 (19.9) | 0.05 | 0.63 |
| Month 8 | 269 | 39 (14.5) | 272 | 53 (19.5) | 0.12 | >0.99 |
| Month 12 | 273 | 41 (15.0) | 273 | 51 (18.7) | 0.25 | 0.86 |
| Month 24 | 265 | 43 (16.2) | 259 | 51 (19.7) | 0.30 | 0.73 |
| Month 36 | 247 | 35 (14.2) | 252 | 49 (19.4) | 0.12 | 0.79 |
| Month 48 | 151 | 25 (16.6) | 136 | 20 (14.7) | 0.67 | 0.21 |
| Month 60 |  30 |  4 (13.3) |  34 |  5 (14.7) | 0.88 | 0.93 |
|  |   |   |   |   |  |   |
| **Overall‡** |   |   |   |   |  |   |
| Ratio of odds ratios | 1.00 |   |   |  |   |
| 95% CI | (0.71, 1.43) |   |   |  |   |
| P-value | 0.98 |   |   |  |   |

† Chi-squared test, comparing the proportions of participants with NCI between the immediate and deferred group.

**§** P-value for comparing the proportions of participants with NCI between the immediate and deferred group at a given visit after adjustment for differences in baseline prevalence of NCI. Prevalence of NCI at baseline and the given follow-up visit is modeled using generalized estimating equations (GEE) for binary outcomes, with an indicator variable for visit (baseline versus the given follow-up visit), treatment group, and the visit by treatment group interaction. The p-value is for the visit by treatment group interaction.

**‡** Comparison of the change in NCI prevalence between the immediate versus deferred groups from baseline throughout follow-up, using a GEE model for binary outcomes. The treatment effect is modelled as a ratio (immediate/deferred) of odds ratios (odds of NCI through follow-up/odds of NCI at baseline within treatment group). Within each treatment group, the odds ratio (follow-up/baseline) models the change in the prevalence of NCI from baseline.

**Table S4.** Changes in the Global Deficit Score (GDS), from baseline through follow-up**.**

|  |
| --- |
|  |
|  |
|  | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred)†** |
|  |  |  |  |
| **Visit** | **N** | **Mean** | **N** | **Mean** | **Difference** | **95% C.I** | **P-value** |
| Month 4 | 273 | -0.39 | 282 | -0.22 | -0.17 | (-0.48, 0.14) | 0.28 |
| Month 8 | 269 | -0.44 | 272 | -0.19 | -0.26 | (-0.59, 0.07) | 0.13 |
| Month 12 | 273 | -0.44 | 273 | -0.54 | 0.10 | (-0.23, 0.42) | 0.56 |
| Month 24 | 265 | -0.55 | 259 | -0.34 | -0.21 | (-0.57, 0.15) | 0.25 |
| Month 36 | 247 | -0.39 | 252 | -0.26 | -0.13 | (-0.48, 0.22) | 0.47 |
| Month 48 | 151 | -0.33 | 136 | -0.40 |  0.08 | (-0.36, 0.51) | 0.73 |
| Month 60 |  30 | -0.67 |  34 |  0.21 | -0.88 | (-1.90, 0.14) | 0.10 |
|  |   |   |   |   |  |  |  |
| Overall‡ |   | -0.44 |   | -0.33 | -0.11 | (-0.35, 0.12) | 0.35 |
| Through Month 12‡ |   | -0.40 |   | -0.32 | -0.08 | (-0.32, 0.17) | 0.53 |

**†** At each visit, the treatment difference was estimated in a linear regression model that included the treatment group indicator and the baseline z-score.

‡ The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline GDS, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S5.** Changes in the Center for Epidemiological Studies Depression scale (CES-D), from baseline through follow-up.

|  |
| --- |
|  |
|  |
|  | **Immediate Group** | **Deferred Group** | **Estimated Difference (Immediate - Deferred)†** |
|  |  |  |  |
| **Visit** | **N** | **Mean** | **N** | **Mean** | **Difference** | **95% CI** | **P-value** |
| Month 4 | 268 | 0.67 | 277 | 1.11 | -0.45 | (-1.90, 1.01) | 0.55 |
| Month 8 | 263 | -0.42 | 268 | 0.24 | -0.66 | (-2.12, 0.79) | 0.37 |
| Month 12 | 268 | -1.08 | 270 | 0.21 | -1.30 | (-2.86, 0.27) | 0.10 |
| Month 24 | 261 | -1.24 | 254 | -0.18 | -1.06 | (-2.61, 0.49) | 0.18 |
| Month 36 | 244 | -1.21 | 249 | -0.49 | -0.71 | (-2.31, 0.89) | 0.38 |
| Month 48 | 145 | -1.01 | 135 | -0.79 | -0.23 | (-2.51, 2.05) | 0.84 |
| Month 60 |  29 | -0.53 |  33 | -1.99 |  1.46 | (-3.20, 6.13) | 0.54 |
|  |   |   |   |   |  |  |  |
| Overall‡ |   | -0.69 |   | -0.10 | -0.59 | (-1.63, 0.45) | 0.27 |
|  |  |  |  |  |  |  |  |
| Through Month 12‡ |   | -0.23 |   | 0.47 | -0.71 | (-1.84, 0.43) | 0.22 |

**†** At each visit, the treatment difference was estimated in a linear regression model that included the treatment group indicator and the baseline z-score.

‡ The overall treatment difference was estimated in a linear mixed model with random intercept that included the following fixed effects: treatment group indicator, baseline CES-D, and the visit number as categorical variable. Through month 12, the treatment difference was estimated in a similar linear mixed model.

**Table S6.** Number and percent of participants with depression, defined by the Center for Epidemiological Studies Depression scale (CES-D) > 16.

|  |  |  |  |
| --- | --- | --- | --- |
|   | **Immediate Group** | **Deferred Group** |   |
| **Visit** | **No. of pts in group** | **N (%) of ptswith depression** | **No. of pts in group** | **N (%) of ptswith depression** | **P-value†** |
| Baseline | 272 | 86 (31.6%) | 288 | 92 (31.9%) |   |
| Month 4 | 264 | 94 (35.6%) | 276 | 101 (36.6%) | 0.81 |
| Month 8 | 259 | 75 (29.0%) | 261 | 87 (33.3%) | 0.28 |
| Month 12 | 266 | 77 (28.9%) | 258 | 94 (36.4%) | 0.07 |
| Month 24 | 259 | 70 (27.0%) | 254 | 82 (32.3%) | 0.19 |
| Month 36 | 235 | 69 (29.4%) | 247 | 72 (29.1%) | 0.96 |
| Month 48 | 140 | 38 (27.1%) | 130 | 45 (34.6%) | 0.18 |
| Month 60 |  26 |  9 (34.6%) |  33 |  8 (24.2%) | 0.39 |
|  |   |   |   |   |   |
| **Overall‡** |   |   |   |   |   |
| Ratio of odds ratios | 0.82 |   |   |   |
| 95% CI | (0.61, 1.11) |   |   |   |
| P-value | 0.21 |   |   |   |
| † Chi-squared test, comparing the proportions of participants with NCI between the immediate and deferred group. |
| **‡** Comparison of the change in the prevalence of depression (CES-D > 16) between the immediate versus deferred groups throughout follow-up, using a generalized estimating equations (GEE) model for binary outcomes. The treatment effect is modelled as a ratio (immediate/deferred) of odds ratios (odds of depression through follow-up/odds of NCI at baseline within treatment group). |

**Table S7.** Selected baseline characteristics by pre-specified ART regimen.

|  |  |  |
| --- | --- | --- |
|  |  | **Median [IQR] or N (%)** |
| **Characteristic** |   | **EFV-Containing ART(n= 466)** | **Other ART(n= 126)** |
| Age, years | 33 [26 - 42] | 35 [30 - 45] |
| Female (%) | 49 (10.5) | 18 (14.3) |
| Race (%) |   |   |
| Black | 70 (15.0) | 20 (15.9) |
| Latino/Hispanic | 83 (17.8) | 13 (10.3) |
| Asian | 91 (19.5) | 4 (3.2) |
| White | 194 (41.6) | 86 (68.3) |
| Other | 28 (6.0) | 3 (2.4) |
| Highest formal training (%) |   |   |
| No formal training | 99 (21.2) | 22 (17.5) |
| Vocational training, completed | 110 (23.6) | 35 (27.8) |
| Some college or university | 113 (24.2) | 29 (23.0) |
| Bachelor's degree or higher | 144 (30.9) | 40 (31.7) |
| Currently employed (%) | 357 (76.6) | 95 (75.4) |
| Urban residence (%) | 408 (87.6) | 111 (88.1) |
| Country of enrolment (%) |   |   |
| Argentina/Chile | 76 (16.3) | 7 (5.6) |
| Brazil | 156 (33.5) | 9 (7.1) |
| European countries1 | 51 (10.9) | 51 (40.5) |
| Thailand | 87 (18.7) | 2 (1.6) |
| United Kingdom/Australia | 45 (9.7) | 21 (16.7) |
| United States | 51 (10.9) | 36 (28.6) |
| Time since HIV diagnosis (years) | 0.7 [0.2 - 2.3] | 1.1 [0.4 - 3.6] |
| Likely mode of HIV infection (%) |   |   |
| Injection drug use | 3 (0.6) | 2 (1.6) |
| Male sexual contact with person of same sex | 353 (75.8) | 89 (70.6) |
| Sexual contact with person of opposite sex | 88 (18.9) | 27 (21.4) |
| Other/unknown | 22 (4.7) | 8 (6.3) |
| CD4 (cells/µL) | 632 [574 - 742] | 626 [581 - 741] |
| Nadir CD4 (cells/µL) | 533 [470 - 637] | 545 [474 - 619] |
| CD4:CD8 ratio | 0.65 [0.47 - 0.84] | 0.61 [0.41 - 0.85] |
| HIV RNA (log10copies/mL) | 4.2 [3.6 - 4.6] | 4.3 [3.8 - 4.7] |
| Cardiovascular risk factors and disease |   |   |
| Current smoker (%) | 161 (34.5) | 53 (42.1) |
| Diabetes (%) | 12 (2.6) | 7 (5.6) |
| Hypertension or BP-lowering drugs (%) | 76 (16.3) | 17 (13.5) |
| Hyperlipidemia or lipid-lowering drugs (%) | 45 (9.7) | 19 (15.1) |
| Body mass index (kg/m2) | 23.7 [21.5 - 26.6] | 24.3 [22.2 - 27.6] |
| Any prior CVD diagnosis3 (%) | 6 (1.3) | 2 (1.6) |
| Predicted 10-year Framingham risk of CVD | 1.7 [0.5 - 4.3] | 2.6 [0.8 - 6.3] |
| Laboratory values |   |   |
| Hematocrit (%) | 43.7 [40.8 - 45.7] | 43.0 [41.2 - 45.0] |
| AST/SGOT (U/L) | 26.0 [21.0 - 32.0] | 24.0 [20.5 - 30.0] |
| ALT/SGPT (U/L) | 25.0 [18.0 - 38.0] | 24.5 [19.5 - 34.0] |
| Hepatitis B or C (%) | 27 (5.8) | 10 (8.0) |
| Alcoholism/other substance dependence (%) | 18 (3.9) | 13 (10.3) |
| Psychiatric diagnosis3 (%) | 23 (4.9) | 26 (20.6) |
| CES-D score4 ≥ 16 (%) | 126 (28.6) | 52 (43.3) |
| Global Deficit Score | 1 [0 - 3] | 1 [0 - 3] |
| Mild impairment5 (%) | 92 (19.7) | 25 (19.8) |
| Moderate impairment5 (%) | 13 (2.8) | 3 (2.4) |
| CPE score6, pre-specified ART regimen | 7 [7 - 7] | 7 [7 - 8] |
|  |  |  |

1 Germany, Italy, Belgium, and Switzerland.

2 History of myocardial infarction, stroke, or coronary revascularization.

3 Major depression, bipolar disorder, schizophrenia, or other psychotic disorder.

4 Center for Epidemiological Studies Depression scale, CES-D >16 denotes depression.

5 Internal z-scores below -1 (for mild impairment) or below -2 (for moderate impairment) for 2 or more of the 6 tested domains.

6 Central nervous system penetration efficacy score