Supplementary Material

[Quality control of the dataset 2](#_Toc14621097)

[Attrition bias test 2](#_Toc14621098)

[AX-CPT task 3](#_Toc14621099)

[AX-CPT versions in EDIPPP and EP 4](#_Toc14621100)

[d’ context in CHR clinical outcomes and healthy control at follow-up 4](#_Toc14621101)

[Exploratory alternative models 5](#_Toc14621102)

[8 prediction Models 5](#_Toc14621103)

[Binary logistic regressions for conversion/non-conversion 6](#_Toc14621104)

[Multinomial logistic regressions for CHR three-group outcomes 7](#_Toc14621105)

[Clinical and social functioning scores 8](#_Toc14621106)

# Quality control of the dataset

## Attrition bias test

No significant demographic differences were observed between excluded and included CHR participants, including age, gender, and education. The majority of CHR participants are Caucasian and from University of California, Davis, Oregon and Maine (Supplementary Table 1, selection bias at baseline). Of note, within 54 participants who were excluded in the current study, 48 participants did not have clinical follow-ups, one had persistent symptoms, and five participants remitted at follow-up. Out of 6 excluded participants, 3 have more than 50% AY errors and 1 has more than 55% AX errors. Similarly, there was no significant demographic difference between participants completed follow-up tasks and those who dropped out. The majority of CHR participants are Caucasian and from UC Davis, Oregon and Maine (Supplementary Table 1, attrition at follow-up).

**Supplementary Table 1 Attrition Bias**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Selection bias at baseline** | | Inclusion | Exclusion | Statisticsa (sig.) |
| age | Mean | 16.57 | 15.74 | 3574.00 (0.167) |
| SD | 3.47 | 3.20 |
| N | 117 | 54 |
| Diagnosis | N (Converter/Non-converter) | 19/98 | 0/54 |  |
|  | N (Persistence/Remission) | 46/52 | 1/5 |
|  | N (no clinical follow-ups) |  | 48 |
| Gender | N (M/F) | 67/50 | 38/16 | 2.68 ( 0.102) |
| Male% | 57.3% | 70.4% |
| Education | Mean | 9.88 | 9.00 | 3597.00 (0.074) |
| SD | 2.92 | 2.31 |
| N | 116 (1 missing) | 53 (1 missing) |
| Race | African American/Black (N/%) | 10 | 8 | 7.31 (0.198) |
| Asian American | 5 | 0 |
| Caucasian | 74 (65.5%) | 32 (62.7%) |
| Native Hawaiian or Other Pacific | 3 | 4 |
| Other | 5 | 3 |
| More than one race | 16 | 4 |
| Missing | 4 | 3 |
| Site | Maine (N/%) | 35 (29.7%) | 12 (22.2%) | 7.22 (0.301) |
| Michigan | 10 | 5 |
| New Mexico | 6 | 1 |
| New York | 5 | 4 |
| Oregon | 17 (14.4%) | 9 (16.7%) |
| UC Davis EDIPPP | 23 (19.5%) | 6 (11.1%) |
| UC Davis EP | 21 (17.9%) | 17 (31.5%) |
| **Attrition bias at follow-up** | | Inclusion | Drop-out | Statisticsa (sig.) |
| age | Mean | 15.43 | 17.05 | 1144.50 (0.083) |
| SD | 2.069 | 3.821 |
| N | 35 | 82 |
| Gender | N (M/F) | 18/17 | 49/33 | 0.70 (0.404) |
| Male% | 51.4% | 59.8% |
| Education | Mean | 9.31 | 10.13 | 1.55 (0.125) |
| SD | 2.349 | 3.117 |
| N | 35 | 81 |
| Race | African American/Black (N/%) | 4 | 6 | 4.48 (0.482) |
| Asian American | 2 | 3 |
| Caucasian | 23 (65.7%) | 51 (65.4%) |
| Native Hawaiian or Other Pacific | 0 | 3 |
| Other | 0 | 5 |
| More than one race | 6 | 10 |
| Site | Maine (N/%) | 11 (31.4%) | 24 (28.9%) | 6.66 (3.53) |
| Michigan | 3 | 7 |
| New Mexico | 0 | 6 |
| New York | 0 | 5 |
| Oregon | 4 | 13 |
| UC Davis EDIPPP | 9 (25.7%) | 14 (16.9%) |
| UC Davis EP | 8 (22.9%) | 13 (15.9%) |

a Mann-Whitney U or χ2 test for nonparametric continuous and categorical data analyses

## AX-CPT task

As described in the main manuscript two versions of AX-CPT diagram using in EDIPPP (left panel) and EP (right panel) studies were illustrated in the supplementary figure 1.



Supplementary Figure 1: Two versions of the AX-CPT task were used in the current study with slightly different inter-stimulus-interval (ISI). EDIPPP used longer ISI version (left panel), while EP used shorter ISI version (right panel).

## AX-CPT versions in EDIPPP and EP

CHR individuals utilized different versions of AX-CPT tasks in EDIPPP and EP respectively, however, no significant differences in performance were observed (Supplementary Table 2).

**Supplementary Table 2 AX-CPT version bias**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | EDIPPP (96) | EP (21) | Mann-Whitney U (sig.) |
| AX Hit\_BL | Median | 0.91 | 0.90 | 885.00 (0.382) |
| Q1 | 0.81 | 0.77 |
| Q3 | 0.95 | 0.95 |
| AX Error\_BL | Median | 0.09 | 0.10 | 1131.00 (0.382) |
| Q1 | 0.05 | 0.05 |
| Q3 | 0.19 | 0.23 |
| AY Error\_BL | Median | 0.26 | 0.30 | 1159.00 (0.283) |
| Q1 | 0.13 | 0.17 |
| Q3 | 0.40 | 0.48 |
| BX Error\_BL | Median | 0.15 | 0.17 | 975.00 (0.814) |
| Q1 | 0.05 | 0.04 |
| Q3 | 0.35 | 0.38 |
| BY Error\_BL | Median | 0.00 | 0.00 | 103570.50 (0.804) |
| Q1 | 0.00 | 0.00 |
| Q3 | 0.08 | 0.13 |
| dprimeC\_BL | Median | 0.68 | 0.72 | 995.00 (0.926) |
| Q1 | 0.49 | 0.39 |
| Q3 | 0.84 | 0.82 |

a BL = Baseline

# d’ context in CHR clinical outcomes and healthy control at follow-up

A repeated-measures ANOVA with one within (time) and one between factor (group) was performed to test whether d’ context change over 12-months differed between CHR-R, CHR-P, and CHR-C patient groups. Results showed that there were no significant differences in d’ context at follow up across the three CHR groups (F (2, 35) = 1.54, p = 0.228). There was also no main effect of time to follow up (F (1, 35) = 0.59, p = 0.446) or an interaction between group and time (F (1, 35) = 0.15, p = 0.864)). However, the small sample size available at follow up (N = 76) may have limited our ability to detect significant differences between groups. With statistical power (1 - β) set at 0.80 and alpha set at 0.05 (Cohen, 1988), power analysis results suggested a necessary sample size of 134 to detect between group differences. Thus, the current negative findings could be attributed to a limited sample size. Indeed, exploratory examination of the data shows the highest d’ context in CHR-R (Baseline: median = 0.77, Q1: 0.56 – Q3: 0.91; Follow-up: median = 0.75, Q1: 0.63 – Q3: 0.92) and slightly increased in CHR-P over time (Baseline: median = 0.59, Q1: 0.44 – Q3: 0.82; Follow-up: median = 0.67, Q1: 0.26 – Q3: 0.93), whereas d’ context drastically decreased in CHR-C (Baseline: median = 0.63, Q1: 0.24 – Q3: 0.74; Follow-up: median = 0.55, Q1: 0.41 – Q3: 0.60) (Supplementary Figure 2).



# Exploratory alternative models

## 8 prediction Models

In the current study, we utilized three risk factors (baseline modified P1+P2 sum score, GF:S decline, and GF:R decline) that are not widely computed in the majority of studies. Therefore, we examined universally accepted measures for substitution, including P1 or P2 modified scores and P total scores at baseline; and GF:R or GF:S measures at baseline. That makes 8 predictive models in total (Supplementary Table 3). Binary logistic regressions or multinomial logistic regressions were performed to discriminate psychosis conversion or distinguish CHR-R from CHR-C and CHR-P. Similar approaches were applied to evaluate the prediction performance of these 8 models via the receiver operating characteristic curve (ROC) and area under the ROC Curve (AUC) (Robin *et al.*, 2011).

**Supplementary Table 3 Eight alternative prediction models**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | | Substitutions | | |
|  | **GF:S original model a** | Baseline modified P1+P2 sum | GF:S decline |  |
| Alternative 1 | P total baseline model | P total at baseline | - |
| Alternative 2 | P1 modified model | P1 modified at baseline | - |
| Alternative 3 | P2 modified model | P2 modified at baseline | - |
| Alternative 4 | Baseline GF:S model | - | Baseline GF:S |
|  | **GF:R original model b** | Baseline modified P1+P2 sum |  | GF:R decline |
| Alternative 5 | P total baseline model | P total at baseline | - |
| Alternative 6 | P1 modified model | P1 modified at baseline | - |
| Alternative 7 | P2 modified model | P2 modified at baseline | - |
| Alternative 8 | Baseline GF:R model | - | Baseline GF:R |

a GF:S d’ context validation model

b GF:S d’ context validation model

## Binary logistic regressions for conversion/non-conversion

Exploratory analyses of alternative prediction models failed to improve prediction accuracy suggesting that five selected risk factors in the current study are optimal for predicting for psychosis conversion Noteworthy, only 2 alternative prediction models reached fair prediction accuracy: the alternative model 4 using GF:S baseline scores revealed AUC of 0.701 with 95% confidence interval ranging from 0.574 to 0.828 (Supplementary Figure 1, left panel, purple lines) and the alternative model 8 using GF:R baseline score revealed AUC of 0.708 with 95% confidence interval ranging from 0.576 to 0.840 (Supplementary Figure 1, right panel, purple lines).



Supplementary Figure 2: All 8 alternative prediction models failed to improved prediction performance compared to the original GF:S or GF:R d’ context validation models (redlines).

## Multinomial logistic regressions for CHR three-group outcomes

Two alternative models (model 2 and model 4) that showed comparable prediction accuracy in the binary logistic regression analyses were utilized to predice CHR three subgroups. The results showed improved predictability of CHR three-group outcomes than discrimination of psychosis conversion, which also supported claim of that refined CHR three subgroups may advance our knowledge of the association between cognitive enhancement and CHR transition to psychosis. Of note, the GF:S d’ context validation model showed identical discrimination accuracy regardless of using GF:S baseline scores (AUC = 0.754, 95% CI: 0.68 - 0.83) or and GF:S deline scores (AUC = 0.757, 95% CI: 0.70 - 0.83) (Supplementary Figure 2, left panel). These results suggested that GF:S baseline score is more sensitive to the prediction of CHR three-group outcomes than psychosis conversion. The alternative explanation may be due to 18 missing GF:S decline scores were imputed with GF:S baseline scores. Moreover, using GF:R baseline scores showed lower discrimination accuracy (AUC = 0.761, 95% CI: 0.70 - 0.83) compared to using GF:R decline scores in the GF:R d’ context validation model (AUC = 0.771, 95% CI: 0.71- 0.84) (Supplementary Figure 2, right panel).



Supplementary Figure 2: All four multinomial logistic regression models showed discrimination improvement of CHR three-group outcomes compared to corresponding binary logistic regression models.

# Clinical and social functioning scores

**Table 4 clinical and social functioning scores at baseline and during the 12-month follow-up**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | CHR-R 1 | CHR-P 2 | CHR-C 3 | Statistics (p)b | Post hoc | Statistics (adjusted p)c |
| **Group Differences at Baseline** | | | | | | | |
| P\_BL | Mean | 9.85 | 12.28 | 13.11 | **7.48 (0.001)\*** | 1 vs 2 | **-2.44 (0.006)\*** |
| N | 52 | 46 | 19 | 1 vs 3 | **-3.26 (0.005)\*** |
| SD | 3.46 | 3.93 | 4.35 | 2 vs 3 | -0.74 (1.000) |
| Modified P1+P2\_BL | Mean | 1.67 | 2.50 | 2.84 | **4.41 (0.014)\*** | 1 vs 2 | -0.83 (0.060) |
| N | 52 | 46 | 19 | 1 vs 3 | **-1.17 (0.039)\*** |
| SD | 1.52 | 1.74 | 2.22 | 2 vs 3 | -0.67 (0.060) |
| N\_BL | Mean | 13.18 | 13.61 | 13.44 | 0.07 (0.936) |  | |
| N | 51 | 46 | 18 |
| SD | 6.04 | 5.66 | 5.97 |
| D\_BL | Mean | 5.06 | 5.35 | 6.56 | 1.63 (0.201) |
| N | 51 | 46 | 18 |
| SD | 2.85 | 2.56 | 4.44 |
| G\_BL | Mean | 10.16 | 9.84 | 9.33 | 0.20 (0.816) |
| N | 51 | 45 | 18 |
| SD | 4.44 | 5.07 | 4.89 |
| GF:S\_BL | Mean | 5.98 | 5.98 | 5.76 | 0.20 (0.820) |
| N | 47 | 42 | 17 |
| SD | 1.24 | 1.28 | 1.35 |
| GF:S decline | Mean | 0.93 | 0.95 | 0.86 | 0.40 (0.961) |
| N | 45 | 40 | 14 |
| SD | 1.009 | 1.085 | 1.167 |
| GF:R\_BL | Mean | 5.77 | 4.95 | 5.53 | 1.54 (0.220) |
| N | 48 | 44 | 19 |
| SD | 2.24 | 2.25 | 2.29 |
| GF:R decline | Mean | 1.46 | 2.03 | 0.79 | 2.43 (0.094) |
| N | 46 | 40 | 14 |
| SD | 1.773 | 2.032 | 1.888 |
| **Group Differences over 12-month Follow-up** | | | | | | | |
| P\_chg | Mean | -7.44 | -3.52 | -3.62 | **11.85 (0.000)\*** | 1 vs 2 | **-3.92 (0.000)\*** |
| N | 45 | 42 | 13 | 1 vs 3 | **-3.83 (0.009)\*** |
| SD | 3.44 | 3.87 | 5.84 | 2 vs 3 | 0.09 (1.000) |
| N\_chg | Mean | -5.18 | -3.19 | -1.33 | 2.23 (0.114) |  | |
| N | 45 | 42 | 12 |
| SD | 6.53 | 6.25 | 4.89 |
| D\_chg | Mean | -3.11 | -1.68 | -1.83 | 2.68 (0.074) |
| N | 45 | 41 | 12 |
| SD | 3.08 | 2.79 | 3.24 |
| G\_chg | Mean | -6.54 | -2.26 | -3.50 | **7.34 (0.001)\*** | 1 vs 2 | **-4.28 (0.001)\*** |
| N | 39 | 39 | 12 | 1 vs 3 | -3.04 (0.206) |
| SD | 5.02 | 5.11 | 4.50 | 2 vs 3 | 1.24 (1.000) |
| GF:S\_chg | Mean | 1.03 | 0.26 | 1.00 | **3.52 (0.034)\*** | 1 vs 2 | **0.77 (0.041)\*** |
| N | 40 | 35 | 13 | 1 vs 3 | 0.03 (1.000) |
| SD | 1.39 | 1.221 | 1.35 | 2 vs 3 | -0.74 (0.259) |
| GF:R\_chg | Mean | 1.00 | 0.49 | -1.00 | **3.72 (0.028)\*** | 1 vs 2 | 0.51 (1.000) |
| N | 40 | 35 | 13 | 1 vs 3 | **2.00 (0.023)\*** |
| SD | 2.36 | 2.13 | 2.52 | 2 vs 3 | 1.49 (0.149) |

a abbreviations: P\_chg/BL = SIPS positive scores change/at baseline; N\_chg/BL = SIPS negative scores change/at baseline; D\_chg/BL = SIPS disorganized scores change/at baseline; G\_chg/BL = SIPS general scores change/at baseline; GAF\_chg/BL = GAF scores change/at baseline; GF:S\_chg/BL = GF:S scores change/at baseline; GF:R\_chg/BL = GF:R scores change/at baseline

b ANOVA for 3 groups parametric tests; Kruskal-Wallis for 3 groups non-parametric tests

c t-test for 2 groups parametric tests; Mann-Whitney U for 2 groups non-parametric tests with Bonferroni corrected p value

\* Significant differences reported for P<.05