Rethinking the Effects of Working Memory Training on Executive Functions in Schizophrenia: A Machine Learning Approach

Tongyi Zhang [1†], Meifang Su[1†], Xiaoning Huo[2], and Xin Zhao[1*]

[1]School of Psychology, Northwest Normal University, Lanzhou, China

[2]The Third People's Hospital of Lanzhou, Lanzhou, China

*Corresponding author: Xin Zhao

E-mail: psyzhaoxin@nwnu.edu.cn

[†]These authors contributed equally to this work.

[Words Count: 5,563]

ABSTRACT

***Background*:** Executive dysfunction in schizophrenia profoundly impairs functional outcomes and remains insufficiently addressed by standard pharmacological treatments. Although computerized cognitive training offers a promising intervention, traditional evaluation methods often fail to capture nuanced improvements along the psychosis-health continuum. This study aims to quantify changes in executive function (EF) profiles following cognitive training and identify candidate causal predictors of treatment response.

***Methods*:** Patients with schizophrenia were randomized into adaptive *N*-back training ($n = 32$), non-adaptive 1-back control ($n = 33$), or treatment-as-usual ($n = 29$) groups. EF was evaluated across working memory, cognitive flexibility, and inhibitory control domains at baseline and post-intervention. A support vector machine classifier, trained on an independent sample (195 schizophrenia patients, 169 controls) and calibrated via Platt scaling, quantified EF profile changes. An integrative causal machine learning framework identified baseline predictors of treatment response.

***Results*:** Adaptive working memory training resulted in significant near-transfer effects on untrained working memory updating tasks and reduced general psychopathology symptoms ($p_{fdr} < 0.05$). The probability of being classified as having a neurotypical EF profile increased substantially in the adaptive training group (from 13.21% at baseline

to 38.79% at follow-up, $p < 0.001$), with these changes correlating with symptom reduction. Working memory maintenance and response inhibition emerged as robust causal predictors of treatment response.

***Conclusions*:** Working memory training induces meaningful shifts in EF profiles in schizophrenia, promoting movement along the psychosis-health continuum toward neurotypical functioning. The classifier-based approach provides a more refined assessment of cognitive gains compared to traditional binary measures, while the causal analysis identifies specific EF domains that predict treatment response.

Key words: Schizophrenia, working memory training, machine learning, Granger causality, executive function, psychosis continuum

Introduction

Schizophrenia is a severe psychiatric disorder characterized by positive symptoms (e.g., hallucinations and delusions), negative symptoms (e.g., diminished emotional expression and social withdrawal), and pervasive cognitive deficits that collectively impair daily functioning and quality of life (McCutcheon et al., 2020). Among the affected cognitive domains, executive function (EF) deficits are particularly detrimental due to their strong correlation with poor functional outcomes, such as impaired social relationships and occupational achievement (Bowie et al., 2006; Kurtz et al., 2008; Puig et al., 2008). EF impairments hinder goal-directed behavior and contribute to aggression, violence, poor treatment adherence, and medication noncompliance, ultimately worsening clinical outcomes (Orellana & Slachevsky, 2013). Therefore, cognitive training targeting EF is a critical therapeutic objective. However, current antipsychotic medications have shown limited efficacy in addressing cognitive impairments (McCutcheon et al., 2023), and some anticholinergic treatments may even exacerbate cognitive dysfunction (Lally & MacCabe, 2015). Due to these limitations, neuroplasticity-based computerized cognitive training (CCT) has emerged as a promising adjunctive intervention.

CCT leverages principles of learning-induced neuroplasticity to restore neuromodulatory processes underlying brain structure, function, and connectivity (Y. Zhang et al., 2024). Meta-analyses have demonstrated small-to-moderate effects of CCT across multiple cognitive domains, such as attention and working memory

(Lejeune et al., 2021; Prikken et al., 2019), attributable to cortical neural plasticity associated with newly acquired perceptual and cognitive skills following CCT interventions (Haut et al., 2010; Mothersill & Donohoe, 2019). Numerous studies have also confirmed the efficacy of CCT in enhancing EF in schizophrenia (Subramaniam et al., 2018). However, traditional CCT investigations typically focus on discrete EF domains rather than examining broader shifts along the "psychosis continuum"—the extent to which patients' EF profiles shift toward those of healthy controls (Berberian et al., 2019; Hanlon et al., 2019; Owen & O'Donovan, 2017). EF comprises multiple dimensions, with a severity continuum ranging from nearly normal to globally impaired in schizophrenia (Raffard & Bayard, 2012). Various EF deficits, including impaired inhibition of automatic responses, reduced cognitive flexibility, and challenges in maintaining or updating goal-related information, align with Miyake et al.'s (2000) influential three-factor model of EF—inhibition control, cognitive flexibility/switching, and working memory (Orellana & Slachevsky, 2013). Conceptualizing the effects of cognitive interventions as binary outcomes oversimplifies this complex neuropsychological process. A more sophisticated approach employs quantitative metrics derived from multivariate machine learning algorithms, providing a nuanced representation of a patient's position along the schizophrenia-healthy control continuum (Haas et al., 2021). Quantifying the degree to which CCT shifts patients' EF profiles toward those of healthy controls may enhance the ability to predict and potentially improve real-world functional outcomes.

Previous meta-analyses have reported heterogeneous effects of cognitive interventions on cognitive performance and daily functioning in schizophrenia (McGurk et al., 2007; Wykes et al., 2011), indicating that individual differences in baseline characteristics may influence intervention efficacy. Recent studies have applied machine learning to capture baseline latent variables that predict cognitive intervention outcomes a priori (Shani et al., 2019). For example, Vladisauskas et al. (2022) trained a support vector classifier that predicted improvement in children following cognitive interventions using baseline individual differences. Others have built models to predict adherence to cognitive training protocols (He et al., 2022; Singh et al., 2022) or functioning in early psychosis patients using baseline cognitive data (Walter et al., 2024). Although machine learning methods excel at modeling high-dimensional, nonlinear relationships, they are essentially associative prediction frameworks that cannot directly identify causal relationships. To address this limitation, researchers have proposed integrative models that combine Granger causality with the advantages of machine learning (Biazoli Jr. et al., 2024; Hofman et al., 2021). Granger causality posits that (i) the cause precedes the effect and (ii) the cause must contain unique information that helps predict the effect (Shojaie & Fox, 2022). Critically, and in contrast to interventional accounts of causality, determining Granger causality depends only on prediction, making it straightforward to integrate into data-driven approaches. By leveraging feature ablation and bootstrap significance testing and constraining the temporal precedence of the cause and the informational uniqueness of

the predictive effect, XGBoost can be used to screen for candidate causal factors associated with symptom alleviation. Further, a series of regression validation analyses can be constructed to examine the robustness of the candidate factors' effects after controlling for potential confounders, ensuring the relative independence and robustness of their causal interpretations. Based on this framework, we can carefully identify candidate causal factors for treatment outcomes following cognitive intervention, aiming to screen for factors that exhibit temporal associations and predictive robustness with the alleviation of psychotic symptoms from observational data, thereby providing priority hypotheses for future cognitive intervention research.

The present study aims to advance the evaluation of CCT interventions in schizophrenia by employing multivariate machine learning to quantify shifts in patients' EF profiles along the psychosis continuum. We applied a previously developed classifier to a new longitudinal intervention sample and evaluated the interventions' efficacy in promoting shifts toward a healthy EF profile (Figure 1). Furthermore, we integrated causal inference with machine learning to identify robust baseline predictors that causally determine these shifts. Through this integrated approach, we aim to advance precision psychiatry and guide the development of targeted cognitive interventions for schizophrenia.

Methods

Ethics approval and consent

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2013. All procedures involving human subjects/patients were approved by the Ethics Committees at the Third People's Hospital of Lanzhou City and Northwest Normal University (Lanzhou, China; ethics approval no. NWNU202405). All participants provided written informed consent after receiving comprehensive information about the study aims, procedures, potential risks, and benefits.

Data collection

Participant recruitment

Two samples were recruited from the Third People's Hospital of Lanzhou City. The cross-sectional discovery sample, details of which can be found in T. Zhang et al. (2024), consisted of 195 patients with schizophrenia and 169 healthy controls. The longitudinal intervention sample comprised 144 patients who had received inpatient treatment within the past two years. Participants in the longitudinal intervention sample were randomly assigned to one of three groups (Supplementary Figure S1): (1) a working memory training group receiving adaptive *N*-back training (n = 48); (2) an active control group receiving non-adaptive 1-back training (n = 48); or (3) a treatment-

as-usual group (n = 48). Schizophrenia diagnoses were established by two resident psychiatrists using the International Classification of Diseases, 10th Revision (ICD-10) criteria (F20.9) and confirmed with the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I). All patients were clinically stable, undergoing consistent treatment, with no anticipated medication changes during the study. Inclusion criteria were: (1) age 18–65 years; (2) capacity to provide informed consent and complete study procedures; and (3) adequate communication abilities. Exclusion criteria included: (1) severe physical illnesses; (2) visual impairments; or (3) adverse drug reactions (details in Supplementary Table S1). Symptom severity was evaluated using the Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1987). Detailed clinical information, such as illness duration and medication regimens, was obtained from electronic medical records.

EF Behavioral tasks

This study was based on the influential model, which subdivides EF into three core dimensions (Figure 3A): inhibitory control, working memory (updating and maintenance), and cognitive flexibility (Friedman & Miyake, 2017). We selected five behavioral tasks of varying complexity to measure these EF dimensions (details in Supplementary Methods) (Zhao et al., 2023): (1) *Running-memory updating* task was used to examine working memory updating (Zhao et al., 2022); (2) *Digit span backward* task was used to measure working memory maintenance (span) (Zhao et al., 2023); (3) *Stroop* task was used to measure interference inhibition (Stroop, 1935); (4) *Go/No-Go*

task was used to measure response inhibition (Gomez et al., 2007); and (5) *Number-switching* task was used to measure cognitive flexibility (Kray et al., 2002). All behavioral tasks were performed using *E*-Prime 3.0 software (Psychology Software Tools, Inc., Pittsburgh, PA, USA).

Computerized training programs

Two computerized working memory training programs were used: an adaptive *N*-back task for the training group and a non-adaptive 1-back task for the active control group.

***Adaptive N-back WM training task*:** The adaptive *N*-back training program included three tasks targeting different domains: animals, spatial locations (Mario), and letters (Zhao et al., 2022). In each task, a series of 5, 7, 9, or 11 stimuli were presented sequentially (Figure 3C). Participants were required to continuously remember the last three items presented. At the end of each series, the nine stimuli were displayed, three of which corresponded to the final three stimuli from the current series. Participants were required to select these three target stimuli in the correct order and received accuracy feedback after each trial. Each training session consisted of six blocks, with each block containing one series of each length (5, 7, 9, and 11 items). The stimulus presentation time started at 1750 ms and was adaptively adjusted based on performance. If accuracy was above threshold, presentation time decreased by 100 ms for the next block. If accuracy fell below threshold, presentation time increased by 100 ms. The presentation time attained at the end of each session was carried over to the beginning

of the next session. Participants in the training group completed all three *N*-back tasks (animals, mario, letters) during each session and were allowed to select the task order themselves. Performance on each task was quantified as the average stimulus presentation time used across the six blocks.

***Non-Adaptive 1-back control training task*:** The active control group completed three non-adaptive 1-back working memory tasks (Figure 3C) identical in design to the adaptive training tasks, except that task difficulty was fixed at the 1-back level (Beloe & Derakshan, 2020). On each trial, participants simply compared the current stimulus with the one presented immediately before it.

Behavioral task and clinical scale analysis

Demographic, clinical, and electronic medical record data were compared between the three participant groups using one-way analysis of variance (ANOVA) for continuous variables and chi-squared tests for categorical variables. False discovery rate (FDR) correction was applied to account for multiple comparisons. EF task performance was analyzed using linear mixed models (LMMs), with group and time point (pre- vs. post-intervention) as fixed effects, and age, sex, education, and olanzapine-equivalent dose as covariates. LMM results were FDR-corrected for multiple comparisons. Correlations between EF measures, PANSS scores, and electronic medical record data were assessed using Pearson's correlation coefficients, with FDR correction.

Evaluating the Continuum of Executive Dysfunction via Machine Learning Classifiers Following Working Memory Training

A previously developed support vector machine (SVM) classifier, trained to discriminate individuals with schizophrenia (SCZ) from healthy controls (HCs) using seven EF assessments across three dimensions (T. Zhang et al., 2024), was applied to the intervention sample at baseline and follow-up, without retraining (Figure 1). The classifier generated subject-specific linear SVM decision scores for each patient at both timepoints. SVM-predicted probabilities were calibrated to align with expected class distributions and clinical reality using Platt scaling, fitting logistic regression to the decision scores of the original HC-SCZ model and applying it to the intervention dataset classifications (Haas et al., 2021). To control for potential confounders, the intervention group's EF scores at both timepoints were normalized using the healthy control group's mean and standard deviation, and adjusted for age, gender, and education level using regression weights estimated from the healthy population. The residualized EF scores (adjusted EF scores after removing covariate effects) were input into the SVM to obtain decision values for each schizophrenia participant at baseline and follow-up. The difference in probability scores between follow-up and baseline, termed "changes in the continuum of executive dysfunction," quantifies the direction of shift across the SVM hyperplane post-intervention. A higher follow-up probability means a shift toward a healthier profile, while a lower probability implies a shift toward a more

psychosis-like profile. Correlation analyses confirmed that the results were not biased by antipsychotic medication or illness duration (Haas et al., 2021).

Integrated Machine Learning and Causal Inference to Predict Executive Dysfunction Continuum Changes

To identify causal factors influencing changes along the executive dysfunction continuum, Granger causality (GC) inference was integrated with XGBoost (Figure 2), followed by univariate and multivariate regression analyses to examine factor robustness while controlling for confounders and establish independent causal contributions. Baseline predictors included demographics, clinical data from electronic medical records, and EF features. The primary outcome was movement along the executive dysfunction continuum, operationalized as the difference in SVM probability scores between baseline and follow-up. An XGBoost model was trained on 70% of the data, with hyperparameters optimized via random three 3-fold cross-validation using *Optuna* (Akiba et al., 2019). Predictive accuracy was validated on the remaining 30% of data (test set) by computing individual-level mean squared error, mean absolute error, and R-squared. Using this comprehensive model, a systematic ablation study was conducted in which each predictor was sequentially excluded to quantify its impact on model performance. Predictors whose exclusion significantly decreased performance ($p < 0.05$; 3,000 FDR-corrected bootstrap iterations) were identified as potential causal factors.

Subsequently, univariate and multivariate regression analyses were performed to examine the robustness of these factors while controlling for confounders and establishing their independent causal contributions. Two sets of multiple regression models were built: one with the full set of predictors, and another including only the potential causal factors obtained by the Granger Causality–XGBoost (GC-XGBoost) method. Additionally, univariate regressions were implemented for each factor independently to obtain uncorrected estimates. If the factors are causally related to the outcomes and independent of each other, the uncorrected and corrected regression coefficients from single and multiple regressions would be similar. Convergent results across these regression models would reinforce that the candidate causal factors are independent of each other (Biazoli Jr. et al., 2024).

Results

Sample Characteristics and Treatment Participation

The cross-sectional discovery sample (N = 364) comprised 195 individuals with schizophrenia (age 35.35 ± 9.35 years, 58.5% male) and 169 HCs (age 37.69 ± 13.71 years, 52.1% male) (Supplementary Table S2, Figure S2). This sample established the baseline EF profiles across the psychosis continuum, as previously reported (Zhang et al., 2024). Subsequently, we recruited 144 eligible patients from the same medical site and randomly assigned them to one of three intervention groups: (1) adaptive working

memory training using adaptive *N*-back tasks ($n = 48$), (2) active control consisting of non-adaptive 1-back training ($n = 48$), or (3) treatment-as-usual (n=48) (Supplementary Figure S1).

Of the 144 participants initially enrolled, 94 (65.3%) completed the trial. In the adaptive working memory training group, 16 participants were lost to follow-up due to inability to continue the intervention ($n = 9$) or premature discharge from the healthcare facility ($n = 7$), resulting in a final sample of 32 participants (age 38.56 ± 10.06 years, 46.88% male). The active control group experienced similar attrition, with 15 participants discontinuing due to inability to complete the intervention ($n = 10$) or premature discharge ($n = 5$), yielding a final sample of 33 participants (age 36.97 ± 9.35 years; 51.52% male). In the treatment-as-usual group, 19 participants did not complete the study protocol due to inability to continue ($n = 9$) or premature discharge ($n = 10$), with 29 participants remaining at study completion (age 36.31 ± 9.16 years; 44.83% male). Importantly, attrition rates did not differ significantly across the three intervention groups ($p = 0.67$). Table 1 presents the demographic and clinical characteristics of participants in each intervention group. As expected for a properly randomized trial, no significant between-group differences were observed at baseline in EF metrics, demographic variables, or clinical characteristics (all $p_{fdr} > 0.05$; Table 1, Supplementary Figure S3).

Computerized Training Outcomes

Training Progress

Figure 3D illustrates the mean stimulus presentation time across the 20 training sessions for the adaptive *N*-back WM training task. A repeated-measures ANOVA revealed significant main effects of training session ($p < 0.001$) for all tasks, indicating consistent reductions in presentation time. Polynomial trend analyses yielded excellent model fits, with significant linear trends for the animal ($\beta = -78.44$, $p = 0.011$), letter, and Mario tasks. Mean presentation times decreased substantially from Session 1 to Session 20 in all tasks (e.g., from 1804.69 ms to 671.88 ms in the animal task). These results suggest that the adaptive *N*-back WM training successfully enhanced WM updating efficiency. Similar improvements were observed in the non-adaptive 1-back WM training task, with significant session effects and robust model fits for all tasks (animal: $R^2 = 0.974$; letter: $R^2 = 0.957$; Mario: $R^2 = 0.982$). Reaction difficulty indices decreased from Session 1 to Session 20 in the animal (from 1.61 to 0.92), letter (1.62 to 0.74), and Mario tasks (1.60 to 0.64).

Near-transfer Effects on Working Memory and General Psychopathology, but No Far-transfer Effects

The results indicate that computerized training elicited near-transfer effects, as evidenced by significant improvements in WM updating performance following adaptive *N*-back WM training. Specifically, in the 1750 ms condition of the WM number-running task, a LMM analysis revealed a significant main effect of time (*F* =

312 3.684, $p < 0.001$) (Figure 4A). Post hoc comparisons indicated that although baseline
313 performance was comparable across groups, the WM training group outperformed both
314 the active control group ($p < 0.001$) and treatment-as-usual group ($p < 0.001$) at follow-
315 up, with no significant difference between the latter two ($p = 0.530$). Parallel findings
316 emerged for the 750 ms condition (Supplementary Table S4).

317 Analysis of clinical outcomes revealed that the PANSS general psychopathology
318 subscale demonstrated a significant group-by-time interaction. LMM analysis indicated
319 a significant main effect of time for general psychopathology ($F = 6.732$, $p < 0.001$)
320 (Figure 4A), with significant reductions in scores from baseline to follow-up observed
321 in all three groups (WM training group: $p < 0.001$; active control group: $p < 0.001$;
322 treatment-as-usual group: $p = 0.006$). Although the overall group effect did not reach
323 significance ($p = 0.549$), post hoc comparisons showed that the WM training group
324 achieved significantly lower general psychopathology scores at follow-up compared
325 with both the active control group ($p = 0.024$) and treatment-as-usual group ($p = 0.022$),
326 with no significant difference between the latter two ($p = 0.442$) (Supplementary Table
327 S4).

328 Notably, computerized adaptive *N*-back WM training did not yield far-transfer
329 effects (Supplementary Figure S5), as training did not lead to improvements in
330 inhibition control, WM maintenance, or cognitive flexibility (other dimensions of EF).
331 Although these functions exhibited significant improvements from baseline to follow-
332 up, no significant between-group differences were observed (all $p > 0.05$).

Working Memory Training Increases Probability of Healthy-Like EF

A SVM classifier accurately distinguished patients with schizophrenia from HCs, achieving a cross-validated AUC of 0.86 (Figure 4B). Residuals from the EF assessments were then included into the SVM classifier, trained on healthy control and schizophrenia data, to generate a predicted probability for each subject, representing their "healthy-like" EF profile. A GLM analysis of these probabilities revealed a significant group-by-time interaction ($p < 0.001$) across the WM training, active control training, and treatment-as-usual groups at baseline and follow-up (Figure 4C). Post hoc Tukey's HSD tests indicated no significant inter-group differences at baseline ($p > 0.05$). However, the WM training group exhibited a substantial increase in the healthy-like probability from 13.21% at baseline to 38.79% at follow-up ($p < 0.001$), while the active control training and treatment-as-usual groups showed no significant changes ($p > 0.05$).

Associations between Healthy-like EF Probability and Clinical Measures

Pearson correlation analyses evaluated the relationship between changes in the predicted healthy-like probability and various baseline cognitive and clinical measures in the overall sample ($N = 94$). Baseline accuracy on the 1750 ms ($r = 0.305$, $p_{fdr} = 0.003$) and 750 ms ($r = 0.249$, $p_{fdr} = 0.016$) conditions of the WM number-running task was significantly positively correlated with the change in predicted probability (Figure 4D). Conversely, baseline reaction time for incongruent stimuli on the Stroop task was

significantly negatively associated with the change in predicted probability (r = –0.245, p_{fdr} = 0.017), indicating that faster responses were linked to a greater increase in healthy-like probability. Baseline accuracy on the *Go/No-Go* task was also positively correlated with the predicted probability change (r = 0.260, p_{fdr} = 0.011).

Regarding clinical symptomatology, significant inverse correlations were observed between the change in predicted probability and changes in PANSS Negative (r = –0.320, p_{fdr} = 0.002), PANSS General Psychopathology (r = –0.335, p_{fdr} = 0.001), and PANSS Total scores (r = –0.320, p_{fdr} = 0.002) (Figure 4D). As expected, the predicted probability was not correlated with olanzapine dosage or illness duration (all p_{fdr} > 0.05), indicating that our results were not biased by antipsychotic medication intake or illness chronicity (Supplementary Figure S6).

Candidate Causal Factors for Predicting Changes in the Continuum of Executive Dysfunction

Identification of Candidate Causal Predictors Using a Granger-XGBoost Framework

To predict changes in the continuum of executive dysfunction, we developed an XGBoost regression model incorporating 24 baseline predictors. The model showed strong performance on the test set (R^2 = 0.45, MSE = 0.04), with over 90% of individual prediction errors below 20% of the maximum possible score (Figure 5A-B). Then, we conducted a systematic ablation study to investigate potential causal predictors, sequentially excluding each baseline predictor from the model over 3,000 bootstrap

resampling iterations. Predictors whose exclusion significantly impaired model performance ($p_{fdr} < 0.05$) were identified as potential causal factors (e.g., smoking status, RM 750 accuracy) (Figure 5C, Supplementary Table S5).

Validation of the Independence of Candidate Causal Factors via Regression Analyses

To validate the independence of candidate predictors identified via the Granger-XGBoost framework, we performed univariate and multivariable regression analyses. Univariate analyses revealed DSBT span and No-go accuracy at baseline as strong predictors of changes in healthy-like probability (both $\beta = 0.121$, $p < 0.001$), with Stroop interference effect also showing a significant relationship ($\beta = 0.069$, $p = 0.006$) (Figure 5D, Supplementary Table S6). RM 750 accuracy had a smaller effect in updated results ($\beta = 0.062$, $p = 0.016$). In a multivariable model using only Granger-XGBoost-identified variables, DSBT span remained the strongest significant predictor ($\beta = 0.099$, $p < 0.001$), followed by No-go accuracy ($\beta = 0.086$, $p = 0.001$) (Figure 5D, Supplementary Table S7). A comprehensive model with all baseline variables (Figure 5D, Supplementary Table S8) confirmed DSBT span ($\beta = 0.079$, $p = 0.003$), No-go accuracy ($\beta = 0.088$, $p = 0.001$), years of education ($\beta = 0.085$, $p = 0.003$), and frequency of episodes ($\beta = 0.053$, $p = 0.024$) as significant predictors of changes in the continuum of executive dysfunction. In summary, our findings indicate that DSBT span (working memory maintenance) and *No-go* accuracy (response control) represent the

most plausible causal candidates for predicting changes in the continuum of executive dysfunction following cognitive intervention.

Discussion

The present study introduces a novel causal machine learning framework to evaluate and predict the efficacy of working memory training in schizophrenia. In contrast to conventional approaches that rely on binary classifications of interventions as simply "effective" or "ineffective," our method quantifies cognitive improvements as continuous shifts along a psychosis-health continuum while identifying candidate causal predictors of these changes. After completing 20 sessions of adaptive *N*-back working memory training, patients with schizophrenia demonstrated significant near-transfer effects to untrained working memory updating tasks compared to both active control and treatment-as-usual groups. These cognitive gains were accompanied by clinically meaningful reductions in general psychopathology symptoms, despite the absence of far-transfer effects to other EF domains (e.g., inhibitory control). Notably, our machine learning classifier revealed a significant shift in EF profiles along the psychosis continuum, with the working memory training group exhibiting a substantial increase in the probability of being classified as having a neurotypical profile, rising from 13.21% at baseline to 38.79% at follow-up ($p < 0.001$). This shift exhibited a significant positive correlation with reductions in PANSS scores. By implementing an integrative model that combines GC with machine learning, we identified working

memory maintenance and response inhibition as the most plausible causal candidates for predicting EF shifts, providing valuable insights for developing personalized cognitive interventions.

Efficacy of Working Memory Training and Transfer Effects

Our engaging working memory training protocol with visually appealing stimuli produced significant near-transfer effects on untrained working memory tasks with similar cognitive demands across 20 sessions. Despite comparable baseline performance across all groups, the working memory training group exhibited superior performance on the *number-running memory* task at follow-up. These findings align with previous research indicating that patients with schizophrenia exhibit neural activation changes similar to HCs and significant cognitive improvements following working memory training (Li et al., 2015) and targeted interventions (Hubacher et al., 2013; Prikken et al., 2019; Subramaniam et al., 2018).

The absence of far-transfer effects to other EF domains (e.g., inhibitory control) is consistent with the process-specific nature of cognitive training effects documented in the literature. Prior research has consistently demonstrated that training specific cognitive processes yield transfer effects primarily on related cognitive constructs rather than enhancing global cognitive functioning (Hubacher et al., 2013; Melby-Lervåg et al., 2016; Minear et al., 2016; Ramsay et al., 2018; von Bastian & Oberauer, 2013; Zhao et al., 2020). Neuroimaging evidence further suggests that working memory training predominantly modulates task-specific neural activity (Cai et al., 2022), with

different training protocols targeting distinct perceptual/cognitive domains and producing distinct neural effects.

Interestingly, we observed a transfer of working memory training benefits to general psychopathology symptoms. Patients who received adaptive working memory training exhibited more substantial improvements in general psychopathology compared to both the active control group and patients receiving standard antipsychotic treatment alone. Specifically, schizophrenia patients exhibited significant reductions in scores on the general psychopathology subscale of the PANSS following working memory training, echoing previous findings (Moritz et al., 2014; Pontes et al., 2013). While standard antipsychotic treatment effectively reduces overall symptom severity in schizophrenia, our results indicate that augmenting standard treatment with working memory training produces enhanced therapeutic outcomes. Previous research has shown that these augmented improvements remain detectable at 6-month follow-up assessments (Fekete et al., 2022). Working memory training appears to effectively ameliorate inattention symptoms and enhance working memory efficiency, potentially reducing cognitive load during daily activities and thereby liberating cognitive resources for improved emotional regulation and social functioning (Li et al., 2019). These improvements are particularly significant given the intrinsic link between inattention symptoms and general psychopathological manifestations, including thought disturbances, emotional dysregulation, and social withdrawal in schizophrenia.

A Novel Framework for Evaluating Cognitive Training Efficacy

A key innovation of our study was the application of machine learning classification to quantify shifts in EF profiles along the psychosis continuum following cognitive control training. Rather than utilizing binary outcome assessments (e.g., "effective" versus "ineffective"), we conceptualized EF changes as shifts along a continuum from psychosis-like toward neurotypical profiles. This approach provides a more nuanced understanding of cognitive improvement following intervention. Our SVM classifier, which achieved an area under the curve of 0.86 in distinguishing patients with schizophrenia from HCs (T. Zhang et al., 2024), revealed that the working memory training group exhibited a substantial increase in the probability of being classified as having a neurotypical EF profile—from 13.21% at baseline to 38.79% at follow-up ($p < 0.001$). Neither the active control group nor the treatment-as-usual group showed significant changes in classification probabilities. These findings suggest that adaptive working memory training not only improves specific EF processes but also promotes a broader reorganization of executive functioning that more closely resembles that of individuals without psychosis. Moreover, the significant correlations observed between changes in the predicted neurotypical probability and reductions in PANSS scores further highlight the clinical relevance of these EF shifts.

While most cognitive remediation studies in schizophrenia have employed univariate analytical techniques (Prikken et al., 2019), such approaches may lack sensitivity to detect subtle, distributed improvements that machine learning algorithms

can identify in high-dimensional data. Our classifier probably captured not only changes in absolute performance levels but also modifications in the pattern of relationships among EF dimensions. Schizophrenia is characterized by disrupted connectivity among cognitive processes (Sheffield & Barch, 2016), and working memory training may have partially normalized these inter-dimensional relationships, resulting in a more integrated EF architecture without necessarily improving each component to the same degree. Enhanced working memory updating ability may have served as a core cognitive process in enhancing functional coordination between different EF components, even when these components showed no measurable improvements when assessed in isolation. This interpretation aligns with mounting evidence that working memory updating may be a fundamental process underlying other cognitive deficits in schizophrenia (Anticevic et al., 2013; Galletly et al., 2007). Furthermore, improvements in this updating function can transfer to daily life situations and various EF tasks (Levaux et al., 2009; Reeder et al., 2004). Our findings extend this work by demonstrating that such improvements are associated with a global shift in the EF profile toward a more neurotypical pattern, suggesting a reorganization of cognitive architecture with significant functional implications.

By leveraging Platt scaling to transform SVM decision scores into clinically interpretable probability values, we enable clinicians to intuitively understand patients' cognitive states and their changes over time. This transformation enhances the interpretability of assessment results and provides a standardized evaluation framework

for cognitive training interventions. This framework quantifies individualized intervention effects and offers a reliable tool for future large-scale, multi-center studies and clinical translation.

Candidate Causal Factors for Predicting Changes in the Continuum of Executive Dysfunction

Previous research has established that baseline cognitive abilities are significant predictors of an individual's response to cognitive interventions (Foster et al., 2017). For patients with schizophrenia, accurately identifying causal relationships among potential factors is critical for optimizing treatment outcomes. While previous studies have employed machine learning algorithms (e.g., LASSO) to predict intervention responses (He et al., 2022; Ramsay et al., 2018; Vladisauskas et al., 2022), these approaches have been limited by their reliance on correlational rather than causal relationships.

We addressed this limitation by implementing an integrative model combining GC with machine learning, enabling identification of candidate causal factors that influence treatment outcomes across the psychosis continuum (Hofman et al., 2021). Our analyses revealed several baseline characteristics that significantly predict response to treatment, including sociodemographic factors (e.g., residence, smoking status) and EF features (e.g., *No-go* accuracy). To determine whether these factors function independently, we conducted regression analyses with causal candidate factors alone as well as all baseline factors as predictors. Performance on the backward digit span

test (working memory maintenance) and accuracy on the *No-go* trials (response inhibition) emerged as relatively independent cognitive predictors, maintaining significance consistently across all regression models.

These findings align with previous research demonstrating that cognitive intervention gains can be predicted by pre-existing cognitive traits (Vladisauskas et al., 2022; Walter et al., 2024), supporting the "Matthew effect" in cognitive training. Previous studies found that *Go/NoGo* task performance predicts completion of substance abuse treatment (Steele et al., 2014), and individuals with high working memory capacity show larger gains from EF training (Foster et al., 2017). Working memory likely facilitates retention of training strategies, while inhibitory control enables patients to maintain focus during intervention sessions. These results have important clinical implications for personalizing cognitive interventions in schizophrenia. Treatment approaches could be tailored based on individual cognitive profiles, with patients displaying lower baseline working memory or inhibitory control potentially benefiting from additional support or modified protocols. Alternatively, preliminary interventions specifically targeting these foundational cognitive abilities may improve subsequent response to comprehensive cognitive remediation.

Limitations and Considerations

Despite these promising results, several limitations warrant consideration. First, similar to previous cognitive training studies in psychiatry (Prikken et al., 2019), our sample size was relatively modest, and we did not conduct extended follow-up

assessments to evaluate the long-term maintenance of intervention effects. The attrition rate (approximately 35%) could potentially introduce selection bias, although our analyses indicated that participants who withdrew did not significantly differ from study completers on key sociodemographic characteristics or electronic medical record parameters (see Supplementary Table S3). Future research should address these methodological limitations through larger, more diverse cohorts and implementation of longitudinal follow-up protocols to determine whether the observed improvements in EF profiles persist over time. Second, while our integrative model combining GC with machine learning has been validated for identifying potential causal factors influencing treatment outcomes in dynamic contexts (Biazoli Jr. et al., 2024; Hofman et al., 2021) and provided valuable guidance for personalized intervention development, it relied on longitudinal observational data and GC assumptions rather than establishing definitive causal relationships. Future investigations should control for additional confounding variables to more rigorously examine the causal baseline features that predict cognitive intervention efficacy. Finally, this observational study involved hospitalized patients receiving various antipsychotic medications. We followed a standard protocol by converting these medications to olanzapine equivalents (Gardner et al., 2010; Leucht et al., 2015) and confirmed no significant baseline differences between the working memory training, active control, and treatment-as-usual groups. Nevertheless, the potential differential effects of specific pharmacological agents on EF performance cannot be entirely ruled out. Future research would benefit from more stringent

medication protocols or stratified analyses to better account for these pharmacological variables.

Conclusion

In summary, this study demonstrates that adaptive working memory training significantly improves domain-specific executive functioning in schizophrenia and yields transferable benefits to general psychopathology symptoms, despite the lack of far-transfer effects to other EF domains. Notably, our multivariate machine learning approach revealed that working memory training facilitates substantial shifts in EF profiles along the psychosis-health continuum, propelling patients toward neurotypical functioning. Moreover, the identification of baseline working memory maintenance capacity and response inhibition ability as the most plausible causal predictors of changes in executive functioning following cognitive intervention offers valuable insights for developing personalized treatment strategies for individuals with schizophrenia.

Authors' contributions. TZ designed the study, conducted the analysis, and drafted the manuscript. MS acquired the data and drafted the manuscript. XH made significant contributions to data acquisition. XZ designed the study, critically revised the manuscript, and provided funding.

Funding Statement: This work was supported by the National Natural Science Foundation of China (No. 32260207 [to Xin Zhao]).

Competing interest. The authors declare no conflicts of interest.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2013. All procedures involving human subjects/patients were approved by the Ethics Committees at the Third People's Hospital of Lanzhou City and Northwest Normal University (Lanzhou, China; ethics approval no. NWNU202405). All participants provided written informed consent after receiving comprehensive information about the study aims, procedures, potential risks, and benefits.

Data availability statement. The raw data of our sample are protected and are not publicly available due to data privacy. These data can be accessed upon reasonable request to the corresponding author (X. Z.). Derived data supporting the findings of this study are available from the corresponding author (X. Z.) upon request.

Code availability statement. Scripts to run the main analyses have been made publicly available and can be accessed at github.com/tyzhang98/ML-PsyExecShift.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperp. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631.

Anticevic, A., Repovs, G., & Barch, D. M. (2013). Working Memory Encoding and Maintenance Deficits in Schizophrenia: Neural Evidence for Activation and Deactivation Abnormalities. *Schizophrenia Bulletin*, *39*(1), 168–178.

Beloe, P., & Derakshan, N. (2020). Adaptive working memory training can reduce anxiety and depression vulnerability in adolescents. *Developmental Science*, *23*(4), e12831.

Berberian, A. A., Gadelha, A., Dias, N. M., Mecca, T. P., Comfort, W. E., Bressan, R. A., & Lacerda, A. T. (2019). Component mechanisms of executive function in schizophrenia and their contribution to functional outcomes. *Brazilian Journal of Psychiatry*, *41*(1), 22–30.

Biazoli Jr., C. E., Sato, J. R., & Pluess, M. (2024). Causal relationships in longitudinal observational data: An integrative modeling approach. *Psychological Methods*, *29*(6), 1132–1147.

Bowie, C. R., Reichenberg, A., Patterson, T. L., Heaton, R. K., & Harvey, P. D. (2006). Determinants of Real-World Functional Performance in Schizophrenia Subjects: Correlations With Cognition, Functional Capacity, and Symptoms. *American Journal of Psychiatry*, *163*(3), 418–425.

Cai, Y., Yang, C., Wang, S., & Xue, G. (2022). The Neural Mechanism Underlying Visual Working Memory Training and Its Limited Transfer Effect. *Journal of Cognitive Neuroscience*, *34*(11), 2082–2099.

Fekete, Z., Vass, E., Balajthy, R., Tana, Ü., Nagy, A. C., Oláh, B., Domján, N., & Kuritárné, I. S. (2022). Efficacy of metacognitive training on symptom severity, neurocognition and social cognition in patients with schizophrenia: A single-blind randomized controlled trial. *Scandinavian Journal of Psychology*, *63*(4), 321–333.

Foster, J. L., Harrison, T. L., Hicks, K. L., Draheim, C., Redick, T. S., & Engle, R. W. (2017). Do the effects of working memory training depend on baseline ability level? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1677–1689.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204.

Galletly, C. A., MacFarlane, A. C., & Clark, C. R. (2007). Impaired updating of working memory in schizophrenia. *International Journal of Psychophysiology*, *63*(3), 265–274.

Gardner, D. M., Murphy, A. L., O'Donnell, H., Centorrino, F., & Baldessarini, R. J. (2010). International Consensus Study of Antipsychotic Dosing. *American Journal of Psychiatry*, *167*(6), 686–693.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*(3), 389.

Haas, S. S., Antonucci, L. A., Wenzel, J., Ruef, A., Biagianti, B., Paolini, M., Rauchmann, B.-S., Weiske, J., Kambeitz, J., Borgwardt, S., Brambilla, P., Meisenzahl, E., Salokangas, R. K. R., Upthegrove, R., Wood, S. J., Koutsouleris, N., & Kambeitz-Ilankovic, L. (2021). A multivariate neuromonitoring approach to neuroplasticity-based computerized cognitive training in recent onset psychosis. *Neuropsychopharmacology*, *46*(4), 828–835.

Hanlon, F. M., Yeo, R. A., Shaff, N. A., Wertz, C. J., Dodd, A. B., Bustillo, J. R., Stromberg, S. F., Lin, D. S., Abrams, S., Liu, J., & Mayer, A. R. (2019). A symptom-based continuum of psychosis explains cognitive and real-world functional deficits better than traditional diagnoses. *Schizophrenia Research*, *208*, 344–352.

Haut, K. M., Lim, K. O., & MacDonald, A. (2010). Prefrontal Cortical Changes Following Cognitive Training in Patients with Chronic Schizophrenia: Effects of Practice, Generalization, and Specificity. *Neuropsychopharmacology*, *35*(9), 1850–1859.

He, Z., Tian, S., Singh, A., Chakraborty, S., Zhang, S., Lustria, M. L. A., Charness, N., Roque, N. A., Harrell, E. R., & Boot, W. R. (2022). A Machine-Learning Based Approach for Predicting Older Adults' Adherence to Technology-Based Cognitive Training. *Information Processing & Management*, *59*(5), 103034.

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., Vazire, S., Vespignani, A., & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595*(7866), 181–188.

Hubacher, M., Weiland, M., Calabrese, P., Stoppe, G., Stöcklin, M., Fischer-Barnicol, D., Opwis, K., & Penner, I.-K. (2013). Working Memory Training in Patients with Chronic Schizophrenia: A Pilot Study. *Psychiatry Journal*, *2013*(1), 154867.

Kray, J., Li, K. Z., & Lindenberger, U. (2002). Age-related changes in task-switching components: The role of task uncertainty. *Brain and Cognition*, *49*(3), 363–381.

Kurtz, M. M., Wexler, B. E., Fujimoto, M., Shagan, D. S., & Seltzer, J. C. (2008). Symptoms versus neurocognition as predictors of change in life skills in schizophrenia after outpatient rehabilitation. *Schizophrenia Research*, *102*(1), 303–311.

Lally, J., & MacCabe, J. H. (2015). Antipsychotic medication in schizophrenia: A review. *British Medical Bulletin*, *114*(1), 169–179.

Lejeune, J. A., Northrop, A., & Kurtz, M. M. (2021). A Meta-analysis of Cognitive Remediation for Schizophrenia: Efficacy and the Role of Participant and Treatment Factors. *Schizophrenia Bulletin*, *47*(4), 997–1006.

Leucht, S., Samara, M., Heres, S., Patel, M. X., Furukawa, T., Cipriani, A., Geddes, J., & Davis, J. M. (2015). Dose Equivalents for Second-Generation Antipsychotic Drugs: The Classical Mean Dose Method. *Schizophrenia Bulletin*, *41*(6), 1397–1402.

Levaux, M.-N., Vezzaro, J., Larøi, F., Offerlin-Meyer, I., Danion, J.-M., & Van der Linden, M. (2009). Cognitive rehabilitation of the updating sub-component of working memory in schizophrenia: A case study. *Neuropsychological Rehabilitation*, *19*(2), 244–273.

Li, X., Chu, M., Lv, Q., Hu, H., Li, Z., Yi, Z., Wang, J., Zhang, J., Lui, S. S. Y., Cheung, E. F. C., Shum, D. H. K., & Chan, R. C. K. (2019). The remediation effects of working memory training in schizophrenia patients with prominent negative symptoms. *Cognitive Neuropsychiatry*, *24*(6), 434–453.

Li, X., Xiao, Y., Zhao, Q., Leung, A. W. W., Cheung, E. F. C., & Chan, R. C. K. (2015). The neuroplastic effect of working memory training in healthy volunteers and patients with schizophrenia: Implications for cognitive rehabilitation. *Neuropsychologia*, *75*, 149–162.

McCutcheon, R. A., Keefe, R. S. E., & McGuire, P. K. (2023). Cognitive impairment in schizophrenia: Aetiology, pathophysiology, and treatment. *Molecular Psychiatry*, 1–17.

McCutcheon, R. A., Reis Marques, T., & Howes, O. D. (2020). Schizophrenia-An Overview. *Jama Psychiatry*, *77*(2), 201–210.

McGurk, S. R., Twamley, E. W., Sitzer, D. I., McHugo, G. J., & Mueser, K. T. (2007). A Meta-Analysis of Cognitive Remediation in Schizophrenia. *American Journal of Psychiatry*, *164*(12), 1791–1802.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of "Far Transfer." *Perspectives on Psychological Science*, *11*(4), 512–534.

Minear, M., Brasher, F., Guerrero, C. B., Brasher, M., Moore, A., & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training: Evidence for near transfer only. *Memory & Cognition*, *44*(7), 1014–1037.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100.

Moritz, S., Veckenstedt, R., Andreou, C., Bohn, F., Hottenrott, B., Leighton, L., Köther, U., Woodward, T. S., Treszl, A., Menon, M., Schneider, B. C., Pfueller, U., & Roesch-Ely, D. (2014). Sustained and "Sleeper" Effects of Group Metacognitive Training for Schizophrenia: A Randomized Clinical Trial. *JAMA Psychiatry*, *71*(10), 1103–1111.

Mothersill, D., & Donohoe, G. (2019). Neural Effects of Cognitive Training in Schizophrenia: A Systematic Review and Activation Likelihood Estimation Meta-analysis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(8), 688–696.

Orellana, G., & Slachevsky, A. (2013). Executive Functioning in Schizophrenia. *Frontiers in Psychiatry*, *4*.

Owen, M. J., & O'Donovan, M. C. (2017). Schizophrenia and the neurodevelopmental continuum:evidence from genomics. *World Psychiatry*, *16*(3), 227–235.

Pontes, L. M. M., Martins, C. B., Napolitano, I. C., Fonseca, J. R., Oliveira, G. M. R., Iso, S. M. K., Menezes, A. K. P. M., Vizzotto, A. D. B., di Sarno, E. S., & Elkis, H. (2013). Cognitive Training for Schizophrenia in Developing Countries: A Pilot Trial in Brazil. *Schizophrenia Research and Treatment*, *2013*(1), 321725.

Prikken, M., Konings, M. J., Lei, W. U., Begemann, M. J. H., & Sommer, I. E. C. (2019). The efficacy of computerized cognitive drill and practice training for patients with a schizophrenia-spectrum disorder: A meta-analysis. *Schizophrenia Research*, *204*, 368–374.

Puig, O., Penadés, R., Gastó, C., Catalán, R., Torres, A., & Salamero, M. (2008). Verbal memory, negative symptomatology and prediction of psychosocial functioning in schizophrenia. *Psychiatry Research*, *158*(1), 11–17.

Raffard, S., & Bayard, S. (2012). Understanding the executive functioning heterogeneity in schizophrenia. *Brain and Cognition*, *79*(1), 60–69.

Ramsay, I. S., Fryer, S., Boos, A., Roach, B. J., Fisher, M., Loewy, R., Vinogradov, S., & Mathalon, D. H. (2018). Response to Targeted Cognitive Training Correlates with Change in Thalamic Volume in a Randomized Trial for Early Schizophrenia. *Neuropsychopharmacology*, *43*(3), 590–597.

Reeder, C., Newton, E., Frangou, S., & Wykes, T. (2004). Which Executive Skills Should We Target to Affect Social Functioning and Symptom Change? A Study of a Cognitive Remediation Therapy Program. *Schizophrenia Bulletin*, *30*(1), 87–100.

Shani, R., Tal, S., Zilcha-Mano, S., & Okon-Singer, H. (2019). Can Machine Learning Approaches Lead Toward Personalized Cognitive Training? *Frontiers in Behavioral Neuroscience*, *13*.

Sheffield, J. M., & Barch, D. M. (2016). Cognition and resting-state functional connectivity in schizophrenia. *Neuroscience & Biobehavioral Reviews*, *61*, 108–120.

Shojaie, A., & Fox, E. B. (2022). Granger Causality: A Review and Recent Advances. *Annual Review of Statistics and Its Application*, *9*(Volume 9, 2022), 289–319.

Singh, A., Chakraborty, S., He, Z., Tian, S., Zhang, S., Lustria, M. L. A., Charness, N., Roque, N. A., Harrell, E. R., & Boot, W. R. (2022). Deep learning-based predictions of older adults' adherence to cognitive training to support training efficacy. *Frontiers in Psychology*, *13*.

Steele, V. R., Fink, B. C., Maurer, J. M., Arbabshirani, M. R., Wilber, C. H., Jaffe, A. J., Sidz, A., Pearlson, G. D., Calhoun, V. D., Clark, V. P., & Kiehl, K. A. (2014). Brain Potentials Measured During a Go/NoGo Task Predict Completion of Substance Abuse Treatment. *Biological Psychiatry*, *76*(1), 75–83.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Subramaniam, K., Gill, J., Fisher, M., Mukherjee, P., Nagarajan, S., & Vinogradov, S. (2018). White matter microstructure predicts cognitive training-induced improvements in attention and executive functioning in schizophrenia. *Schizophrenia Research*, *193*, 276–283.

Vladisauskas, M., Belloli, L. M. L., Fernández Slezak, D., & Goldin, A. P. (2022). A Machine Learning Approach to Personalize Computerized Cognitive Training Interventions. *Frontiers in Artificial Intelligence*, *5*.

von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language*, *69*(1), 36–58.

Walter, N., Wenzel, J., Haas, S. S., Squarcina, L., Bonivento, C., Ruef, A., Dwyer, D., Lichtenstein, T., Bastrük, Ö., Stainton, A., Antonucci, L. A., Brambilla, P., Wood, S. J., Upthegrove, R., Borgwardt, S., Lencer, R., Meisenzahl, E., Salokangas, R. K. R., Pantelis, C., … Kambeitz-Ilankovic, L. (2024). A multivariate cognitive approach to predict social functioning in recent onset psychosis in response to computerized cognitive training. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *128*, 110864.

Wykes, T., Huddy, V., Cellard, C., McGurk, S. R., & Czobor, P. (2011). A Meta-Analysis of Cognitive Remediation for Schizophrenia: Methodology and Effect Sizes. *American Journal of Psychiatry*, *168*(5), 472–485.

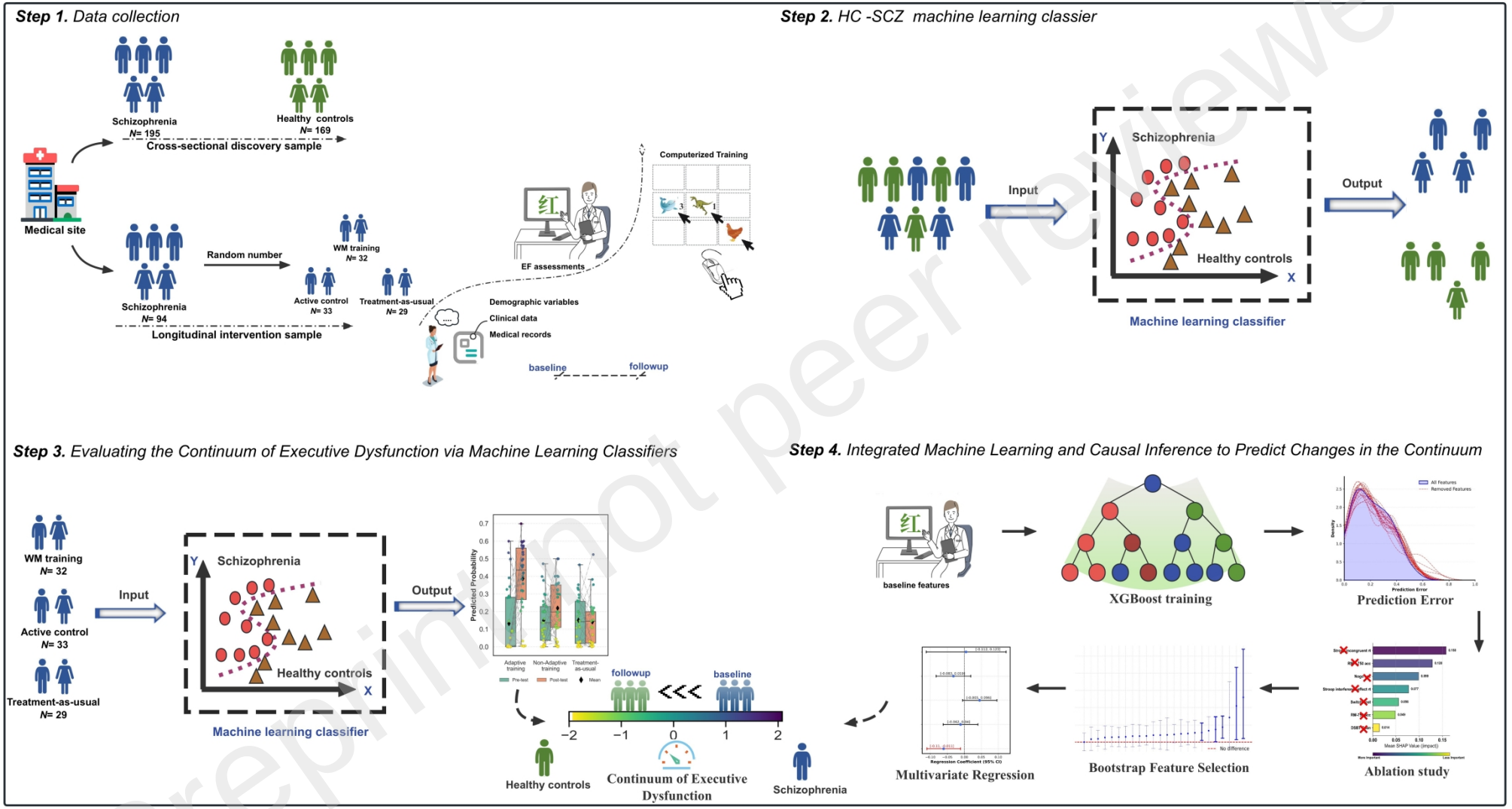
Zhang, T., Zhao, X., Yeo, B. T. T., Huo, X., Eickhoff, S. B., & Chen, J. (2024). Leveraging Stacked Classifiers for Multi-task Executive Function in Schizophrenia Yields Diagnostic and Prognostic Insights. *medRxiv*.

Zhang, Y., Fu, J., & Zhao, X. (2024). Neural correlates of working memory training: An fMRI meta-analysis. *NeuroImage*, *301*, 120885.

Zhao, X., Wang, L., & Maes, J. H. R. (2020). Training and transfer effects of working memory training in male abstinent long-term heroin users. *Addictive Behaviors Reports*, *12*, 100310.

Zhao, X., Wang, Y., & Maes, J. H. R. (2022). The effect of working memory capacity and training on intertemporal decision making in children from low-socioeconomic-status families. *Journal of Experimental Child Psychology*, *216*, 105347.

766 Zhao, X., Zhang, W., Tong, D., & Maes, J. H. R. (2023). Creative thinking and executive functions:
767 Associations and training effects in adolescents. *Psychology of Aesthetics, Creativity, and the*
768 *Arts*, *17*(1), 79–90.



770 **Figure 1. Schematic diagram of the study design and analysis pipelines**

Note: The study consisted of four main steps: **Step 1**: Participant recruitment and assessment. Two samples were recruited: a cross-sectional discovery sample and a longitudinal intervention sample. Participants in both samples underwent identical clinical assessments and completed five paradigms measuring EF across three dimensions. In the longitudinal intervention sample, participants were randomly assigned to one of three groups: 1) a working memory training group receiving adaptive *N*-back training, 2) an active control group receiving non-adaptive 1-back training, or 3) a treatment-as-usual group. **Step 2**: Development of a support vector machine (SVM) classifier. Using the cross-sectional discovery sample data, an SVM classifier was trained to discriminate between individuals with schizophrenia and healthy controls based on seven EF assessments spanning three dimensions measured by five paradigms (Zhang et al., 2024). **Step 3**: Application of the SVM classifier to the longitudinal intervention sample. EF features from the three intervention groups were normalized, adjusted for age, gender, and education level, and then inserted into the SVM classifier developed in Step 2. Platt scaling was used to fit a logistic regression model to the decision scores of the original HC-SCZ model and apply it to the intervention dataset classifications. **Step 4**: Identification of causal factors influencing changes in executive dysfunction. Granger causality (GC) inference was integrated with XGBoost, followed by univariate and multivariate regression analyses, to identify causal factors influencing changes in the continuum of executive dysfunction.

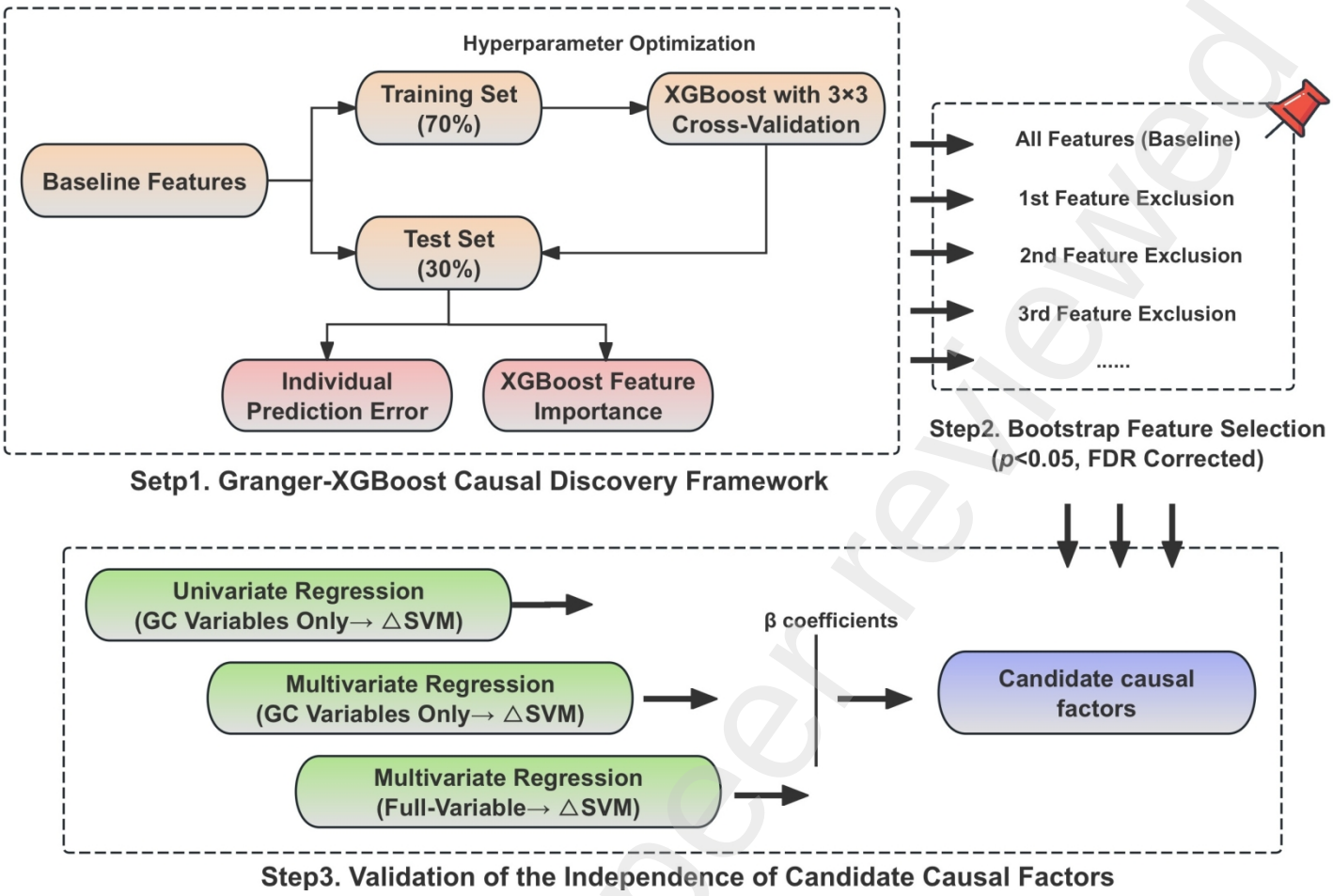
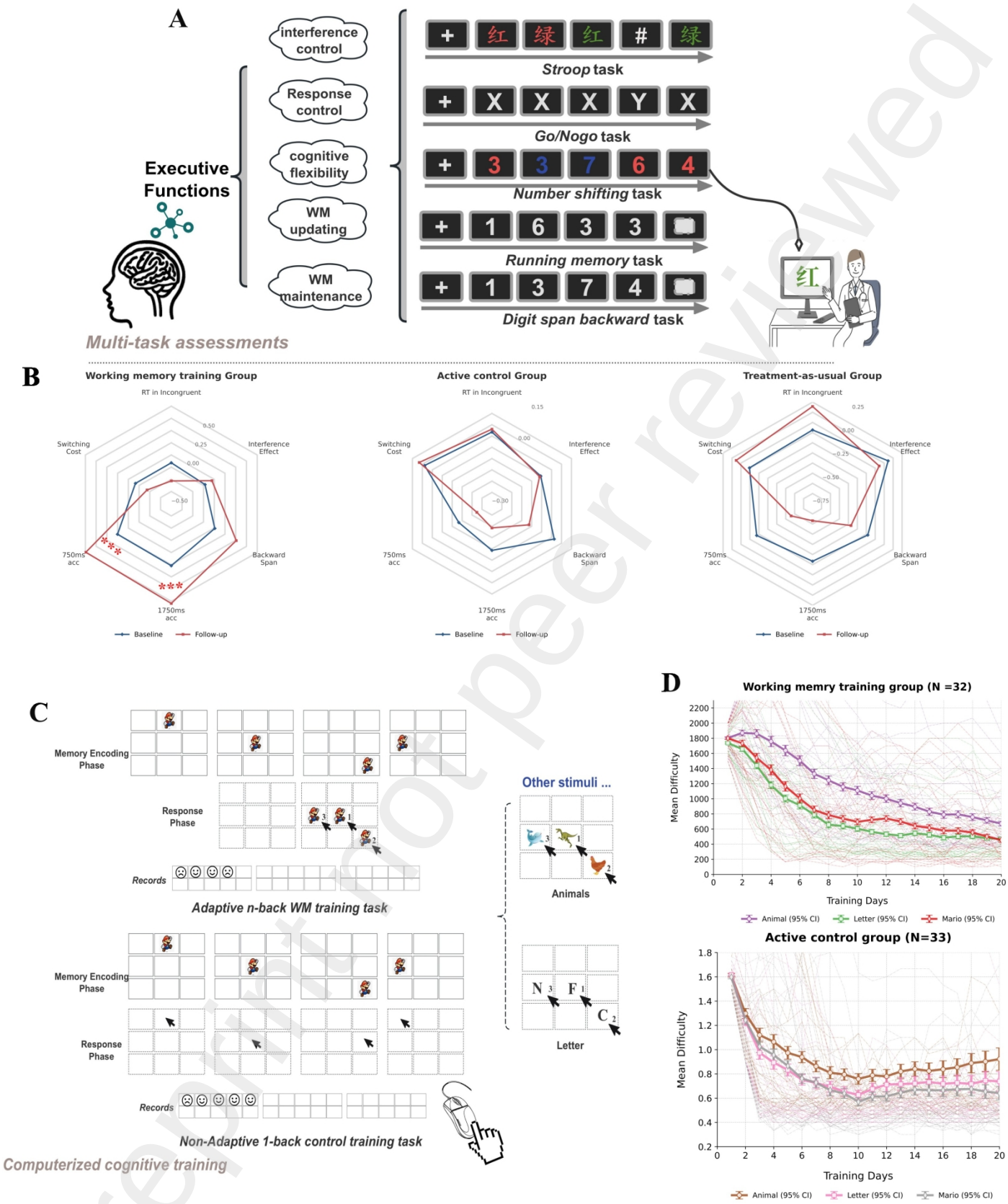
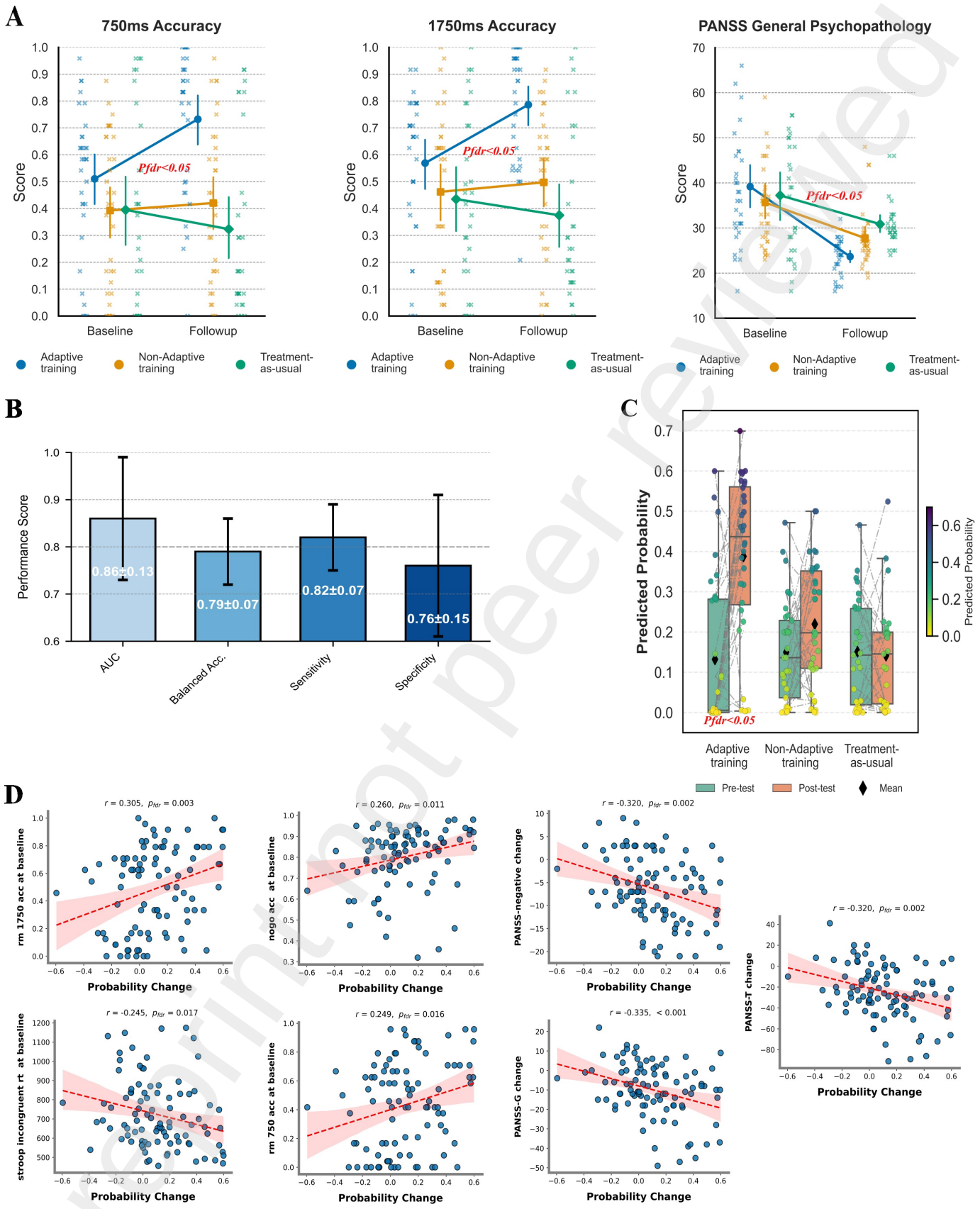


Figure 2. Workflow of Integrated Machine Learning and Causal Inference to Predict Changes in the Continuum of Executive Dysfunction



785 **Figure 3. EF Behavioral Tasks and Working Memory Training**
786 **Procedures**

Note: a) Schematic illustrations of the five behavioral tasks used to measure EF dimensions across varying levels of complexity. **b)** Radar plot comparing EF feature profiles at baseline and follow-up for the three intervention groups: 1) a working memory training group receiving adaptive *N*-back training, 2) an active control group receiving non-adaptive 1-back training, and 3) a treatment-as-usual group. **c)** Schematic diagrams of the computerized training programs: Adaptive *N*-back WM training task and non-adaptive 1-back control training task. Each task included three types of stimuli: animals, spatial locations (Mario), and letters. **d)** Training session schedules for the adaptive *N*-back WM training task and the non-adaptive 1-back WM training task.



795 **Figure 4. Evaluating the Continuum of Executive Dysfunction via**
796 **Machine Learning Classifiers Following Cognitive Intervention**

Note: a) Linear mixed model results examining whether working memory training significantly improved EF relative to an active control group receiving non-adaptive 1-back training and a treatment-as-usual group, with age, sex, education, and olanzapine-equivalent dose as covariates. **b)** Test set classification performance of the SVM classifier for discriminating schizophrenia from healthy controls. **c)** The continuum of executive dysfunction scores for the three intervention groups, obtained by inputting baseline and follow-up EF features into the SVM classifier. **d)** Correlations between change scores in the continuum of executive dysfunction and baseline EF features, as well as change scores in the PANSS ($N = 94$).

Abbreviations: PANSS, Positive and Negative Syndrome Scale; AUC, area under the curve; acc, accuracy; rt, reaction time.

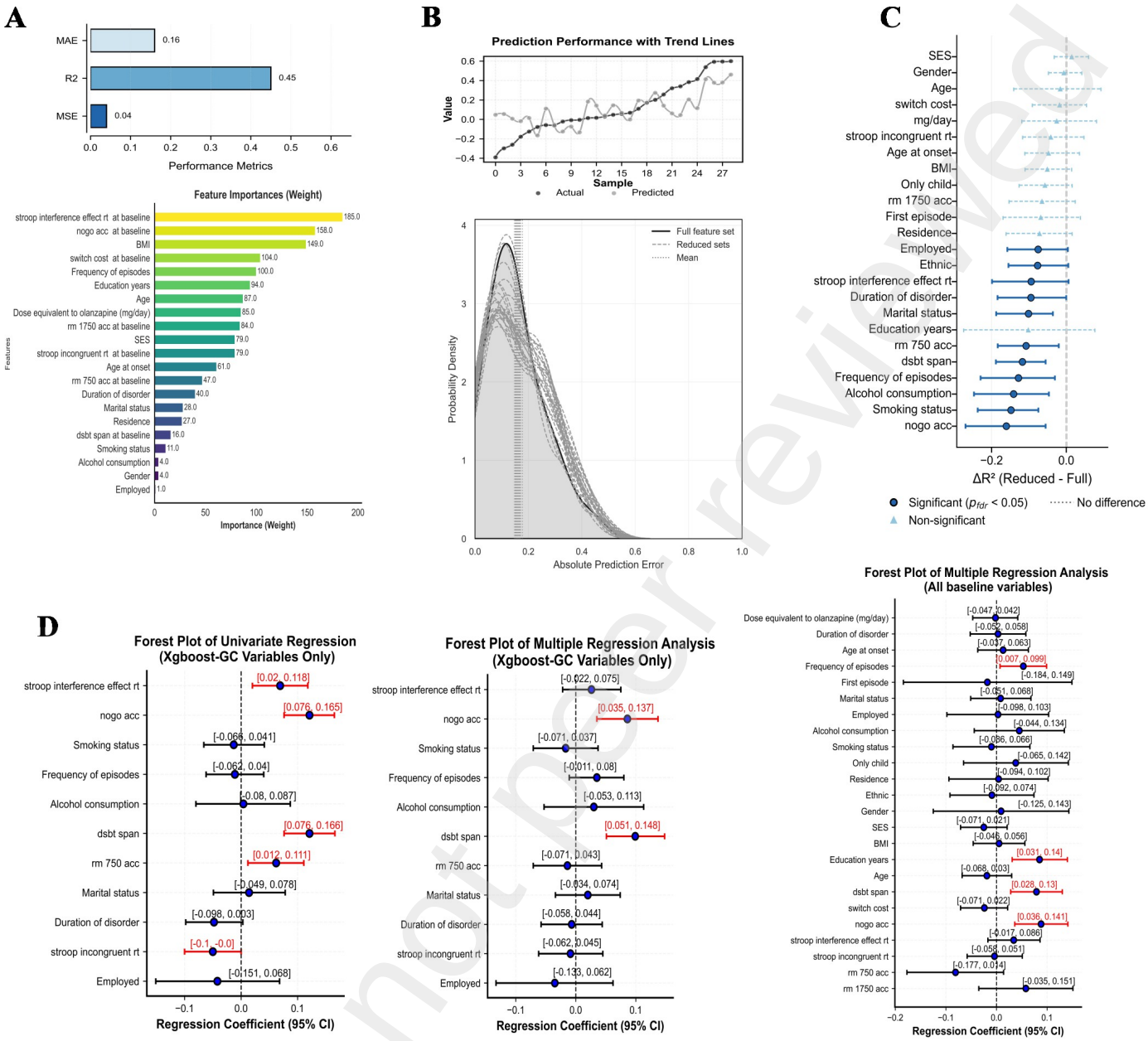


Figure 5. Integrated Machine Learning and Causal Inference to Predict Changes in the Continuum of Executive Dysfunction

Note: a) Performance (top) and feature importance (bottom) of the XGBoost model for predicting changes in the continuum of executive dysfunction using all baseline predictors, including demographics, electronic medical records data, and EF features, evaluated on the test set. **b)** Individual prediction errors for the baseline model (top) and individual prediction errors for the ablation study (bottom). **c)** Predictors whose exclusion significantly decreased performance ($p <$ 0.05; 3,000 false discovery rate-corrected bootstrap iterations) were identified as potential causal factors. **d)** To validate the independence of candidate predictors identified through the Granger-XGBoost framework, univariate and multivariable regression analyses were conducted. (Left) Univariate regression models were run with the full set of predictors. (Middle) Multiple regression

models were run: one with the full set of predictors. (Right) Multiple regression models were run with the full set of predictors.

Abbreviations: MAE, mean absolute error; MSE, mean squared error; BMI, body mass index; SES, socioeconomic status; PANSS, Positive and Negative Syndrome Scale; acc, accuracy; rt, reaction time.

823 **Table 1. Participant demographics and clinical characteristics (*N* = 94)**

(M ± SD)/*n* (%)	WM training group (*n* = 32)	Active control group (*n* = 33)	Treatment-as-usual group (*n* = 29)	$F/\chi 2$	P_{fdr}
Demographic variables					
Age	38.56 ± 10.06	36.97 ± 9.35	36.31 ± 9.16	0.456	0.635
Sex, *n* (%)					
Male	15 (46.88%)	17 (51.52%)	13 (44.83%)	0.296	0.862
Female	17 (53.12%)	16 (48.48%)	16 (55.17%)		
Ethnicity, han, *n* (%)	28 (87.50%)	29 (87.88%)	24 (82.76%)	0.412	0.814
Education, years	11.56 ± 3.83	10.76 ± 3.67	11.34 ± 3.87	0.386	0.681
BMI	23.69 ± 5.33	23.78 ± 3.73	23.43 ± 3.93	0.052	0.949
Residence, urban, *n* (%)	19 (59.38%)	19 (57.58%)	17 (58.62%)	0.022	0.989
SES	23.41 ± 6.05	23.58 ± 6.95	22.10 ± 5.78	0.493	0.612
Employed, *n* (%)	10 (31.25%)	11 (33.33%)	9 (31.03%)	0.047	0.977
Only child, *n* (%)	8 (25.00%)	8 (24.24%)	8 (27.59%)	0.098	0.952
Marital status, *n* (%)					
Unmarried	13 (40.62%)	17 (51.52%)	12 (41.38%)	4.249	0.643
Married	11 (34.38%)	11 (33.33%)	8 (27.59%)		
Divorced	8 (25.00%)	5 (15.15%)	8 (27.59%)		
Widowed	0 (0.00%)	0 (0.00%)	1 (3.45%)		
Smoking history, *n* (%)					
Never	20 (62.50%)	17 (51.52%)	16 (55.17%)	2.619	0.623
1–3 years	2 (6.25%)	1 (3.03%)	3 (10.34%)		
>3 years	10 (31.25%)	15 (45.45%)	10 (34.48%)		

Alcohol consumption history, *n* (%)					
Never	20 (62.50%)	20 (60.61%)	18 (62.07%)		
Occasionally	10 (31.25%)	11 (33.33%)	9 (31.03%)	0.061	0.999
Regularly	2 (6.25%)	2 (6.06%)	2 (6.90%)		
Executive functions					
Interference inhibition (*Stroop* task)					
Reaction times in incongruent stimuli	724.08 ± 196.03	726.57 ± 177.77	719.11 ± 169.69	0.013	0.989
Stroop interference effect	−60.31 ± 65.98	−62.42 ± 94.49	−50.35 ± 81.27	0.189	0.989
Response inhibition (*Go/No-go* task)					
No-Go Accuracy	0.85 ± 0.10	0.79 ± 0.14	0.82 ± 0.13	1.538	0.587
Working memory updating (*Running memory* task)					
Accuracy in 1,750 ms	0.57 ± 0.27	0.46 ± 0.30	0.44 ± 0.32	1.766	0.587
Accuracy in 750 ms	0.51 ± 0.27	0.39 ± 0.29	0.40 ± 0.35	1.550	0.587
Working memory maintenance (*Digit span backward* task)					
Span in digit span backward task	5.34 ± 1.36	5.18 ± 1.65	4.93 ± 2.12	0.441	0.987
Cognitive flexibility (*Number switching* task)					
Switching cost	257.61 ± 233.15	268.01 ± 298.30	245.37 ± 289.92	0.052	0.987
Electronic medication records					
First episode, yes	28 (87.50%)	29 (87.88%)	25 (86.21%)	0.042	0.979
Frequency of episodes	4.75 ± 2.71	4.94 ± 3.31	4.83 ± 3.38	0.030	0.971
Age at onset	26.78 ± 6.24	28.21 ± 7.97	27.28 ± 9.50	0.271	0.763
Duration of disorder	12.02 ± 7.92	11.70 ± 8.48	12.84 ± 9.72	0.138	0.871
Dose equivalent to olanzapine (mg/day)	12.59 ± 6.49	13.13 ± 6.07	13.24 ± 5.94	0.097	0.908

PANSS subscales					
PANSS Positive	19.25 ± 6.29	21.30 ± 6.09	19.31 ± 7.90	0.961	0.387
PANSS Negative	20.97 ± 6.99	20.33 ± 5.90	18.76 ± 6.64	0.919	0.403
PANSS General	39.19 ± 13.68	35.67 ± 12.01	37.24 ± 14.82	0.554	0.576
PANSS Total scores	79.41 ± 25.36	77.30 ± 21.09	75.31 ± 26.77	0.215	0.807

Note: Data are presented as the mean ± standard deviation or n (%). The *p*-values in bold face indicate statistically significant differences ($p < 0.05$). **Calculation:** BMI is calculated as weight (kg) divided by height squared (m^2). **Abbreviations:** BMI, body mass index; SES, socioeconomic status; PANSS, Positive and Negative Syndrome Scale; fdr, false discovery rate correction.

824