Supplemental File Table of Contents

Supplemental Methods	Pages 2 – 5
Supplemental Methods References	Page 6
Supplemental Figure 1 (Consort Diagram)	Page 7
Supplemental Figure 2 (VR-JIT Interface)	Page 8
Supplemental Figure 3 (Intent-to-Treat, Chi-Square Evaluating Employment)	Page 9
Supplemental Figure 4 (Intent to Treat, Cumulative Hazard of Employment)	Page 10
Supplemental Figure 5 (Recent IPS Enrollees, Cumulative Hazard of Employment)	Page 11
Supplemental Table 1 (IPS nonresponders and Recent IPS Enrollees, Logistic Regressions)	Page 12
Supplemental Table 2 (IPS nonresponders and Recent IPS Enrollees, Cox proportional hazard regressions)	Page 13
Supplemental Table 3 (Intent-to-treat, Pearson correlation table of covariates covariates)	Page 14
Supplemental Table 4 (IPS nonresponders and recent IPS enrollees, Pearson correlation tables)	Page 15

Supplemental Methods

Study Measures

Job Interview Skills

The Mock Interview Rating Scale (MIRS) role-plays required participants to complete a job application and select two fictional job openings from a list of eight jobs before engaging in the role-plays (with role-play interviewers trained to fidelity). The role-plays were assessed across 8 job interview skills or items on a five-point scale from 1=poor to 5=excellent. An example of the anchoring system for the skill of 'coming across as a hard worker' reflects a rating of a '1' for participants who make statements about "showing up late," "missing work frequently," and "avoiding responsibilities."

Although the original MIRS consisted of nine items, we excluded the item targeting 'negotiation' where participants were instructed to 'ask for Thursdays off' during the role-play. Fewer than 40% of participants remembered to ask for Thursdays off during their role-plays. This approach is consistent with the efficacy studies of VR-JIT (1-5). Thus, this item was removed from the MIRS assessment for all future studies. Additionally, the MIRS interview role-play includes 13 required job interview questions plus three additional questions (randomly selected from a list of 90 optional questions). Four primary raters were trained by the principal investigator (PI) using 10 gold standard mock interview videos, and then their reliability was monitored to prevent drift by jointly scoring approximately 10% of all videos (ICC=.93). Three additional staff trained on the 10 gold standard videos and rated 38% of the reliability videos as they departed the project before completion: rater 5 (ICC=.99), rater 6 (ICC=.98), and rater 7 (ICC=.91).

Social Competence

The Social Skills Performance Assessment (SSPA) requires participants to complete a brief practice session and then role-play meeting a new neighbor and making a request from a landlord. Each scene was rated on a 5-point scale using anchors across eight criteria for the new neighbor scene (i.e., interest/disinterest, fluency, clarity, focus, affect, social appropriateness, grooming, overall conversation) and nine criteria for the landlord scene (i.e., interest/disinterest, fluency, clarity, focus, affect, social appropriateness, negotiation ability, submissive-persistent, overall argument). Example anchors include: "1=very disinterested. Impaired normal conversation. Asks virtually no questions; gives brief responses. Minimal initiation; passive listener." We computed the mean item-level score for each scene and then computed the mean between scene-level means to reflect a single mean of social competence at pre-test and post-test.

Five primary raters were trained using the SSPA training protocol (6). Their coding reliability was monitored to prevent drift by jointly scoring approximately 10% of all videos (ICC=.97). One additional staff rated 50% of the available reliability videos prior to departing the project (ICC=.99).

Study Procedures

Recruitment

Employment specialists reviewed their caseload for potential participants, provided them with a study flyer, and referred them to either call the study coordinator or attend a research staff-led group presentation to learn about the study. Employment specialists also obtained verbal consent from interested participants for research staff to call and tell them more about the study. Potential participants were screened over the phone or after the presentation. Individuals

who passed the screener were later contacted to schedule two visits to provide informed consent and complete pre-test assessments.

VR-JIT Implementation

To facilitate hierarchical learning using lessons learned from the VR-JIT efficacy trials, study participants were asked to progress from easy to medium to hard virtual interviews. The recommended progression was visualized in a curriculum tracking form that VR-JIT implementers showed to participants and tracked with them. The progression is as follows. First, participants were required to complete at least three interviews on easy. If they achieved a score of 90 or higher on any of the first three 'easy' interviews, then they advanced to the 'medium' difficulty interview. If they did not score 90 or higher on their first three interviews then they had two more attempts to achieve 90 or higher. Participants then automatically advanced to 'medium' after completing five 'easy' interviews, regardless of score. This same model was followed to progress from 'medium' to 'hard' interviews. Then participants were asked to perform 'hard' interviews for the remainder of their training. During training, VR-JIT implementers were instructed that this recommended curriculum was flexible and could be adapted. For example, a participant who completed two interviews on easy and score 90 or better on both could potentially move straight to medium interviews without completing the recommended minimum of three interviews on easy before progressing to medium.

Data Analysis

Power Analysis

Prior efficacy studies of VR-JIT revealed medium-to-large effect-sized improvements in employment (e.g., OR=8.7), which we used to power the study. This is described in detail in our study protocol (7). A post hoc power analysis of our revised sample size (n=90) revealed >=85%

power for a significant (α=.05) two-sided test if the IPS+VR-JIT group has just over twice the employment rate of the IPS group by follow-up (e.g., 59% vs. 25%, or an OR = 4.3).

Missing Data

As noted in the manuscript, n=17 participants did not attend the post-test research assessment visit. Furthermore, n=10 additional participants did not complete the SSPA post-test due to the coronavirus pandemic or their geographic relocation. These extenuating circumstances required a remotely conducted post-test visit via a telephone call. The remote post-test visit required the study team to reduce the number of collected assessments to limit the burden on participants. Lastly, n=3 participants refused to complete the SSPA at the pre-test assessment visit.

References

1. Smith MJ, Ginger EJ, Wright K, et al.: Virtual reality job interview training in adults with autism spectrum disorder. J Autism Dev Disord 44:2450-63, 2014

2. Smith MJ, Ginger EJ, Wright M, et al.: Virtual reality job interview training for individuals with psychiatric disabilities. J Nerv Ment Dis 202:659-67, 2014

3. Smith MJ, Boteler Humm L, Fleming MF, et al.: Virtual Reality Job Interview Training For Veterans with Posttraumatic Stress Disorder. J Vocat Rehabil 42:271-9, 2015

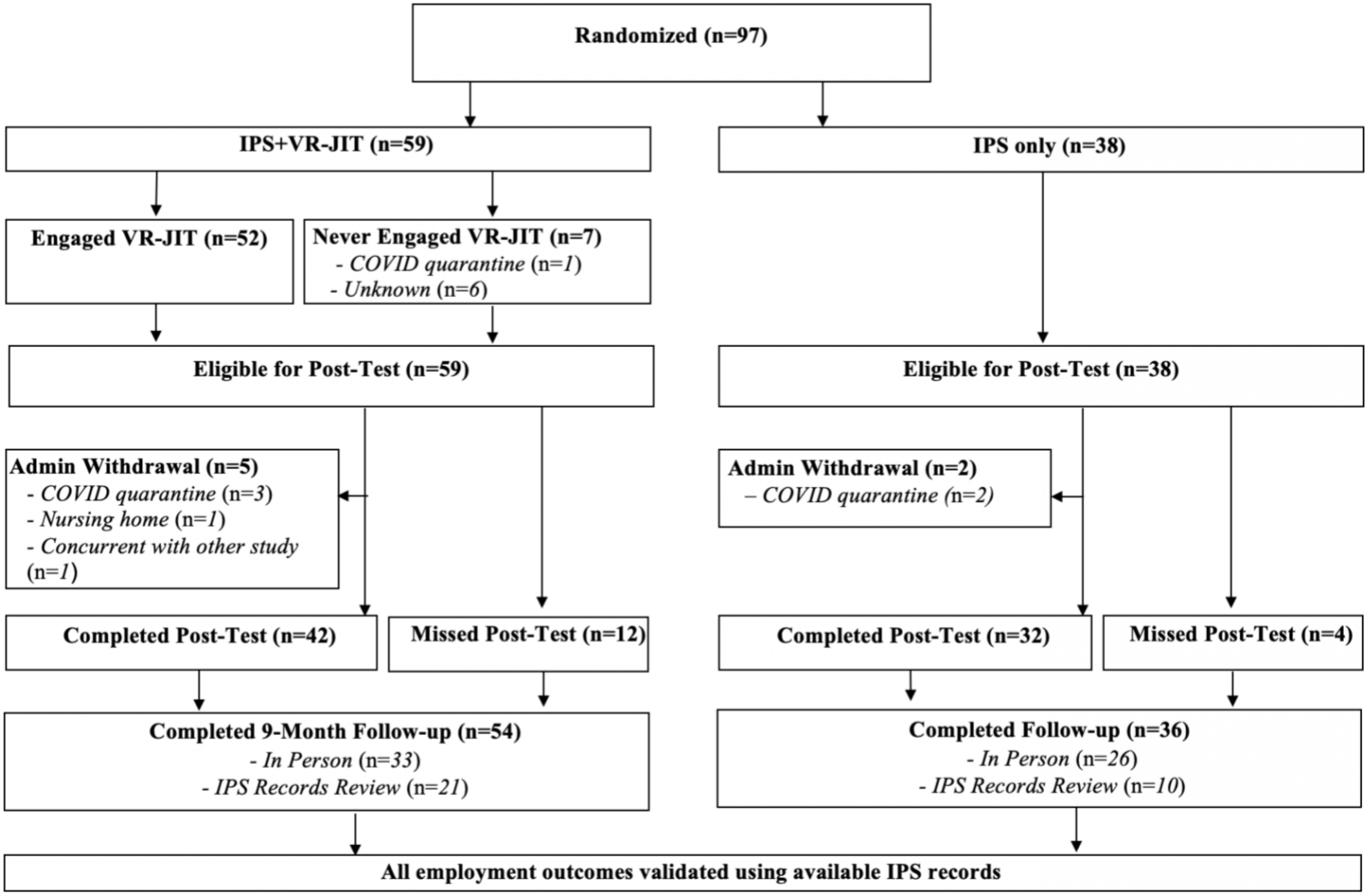
4. Smith MJ, Fleming MF, Wright MA, et al.: Job Offers to Individuals With Severe Mental Illness After Participation in Virtual Reality Job Interview Training. Psych serv 66:1173-9, 2015

5. Smith MJ, Fleming MF, Wright MA, et al.: Virtual reality job interview training and 6-month employment outcomes for individuals with schizophrenia seeking employment. Schizophr Res 166:86-91, 2015

6. Patterson TL, Moscona S, McKibbin CL, et al.: Social skills performance assessment among older patients with schizophrenia. Schizophrenia Research 48:351-60, 2001

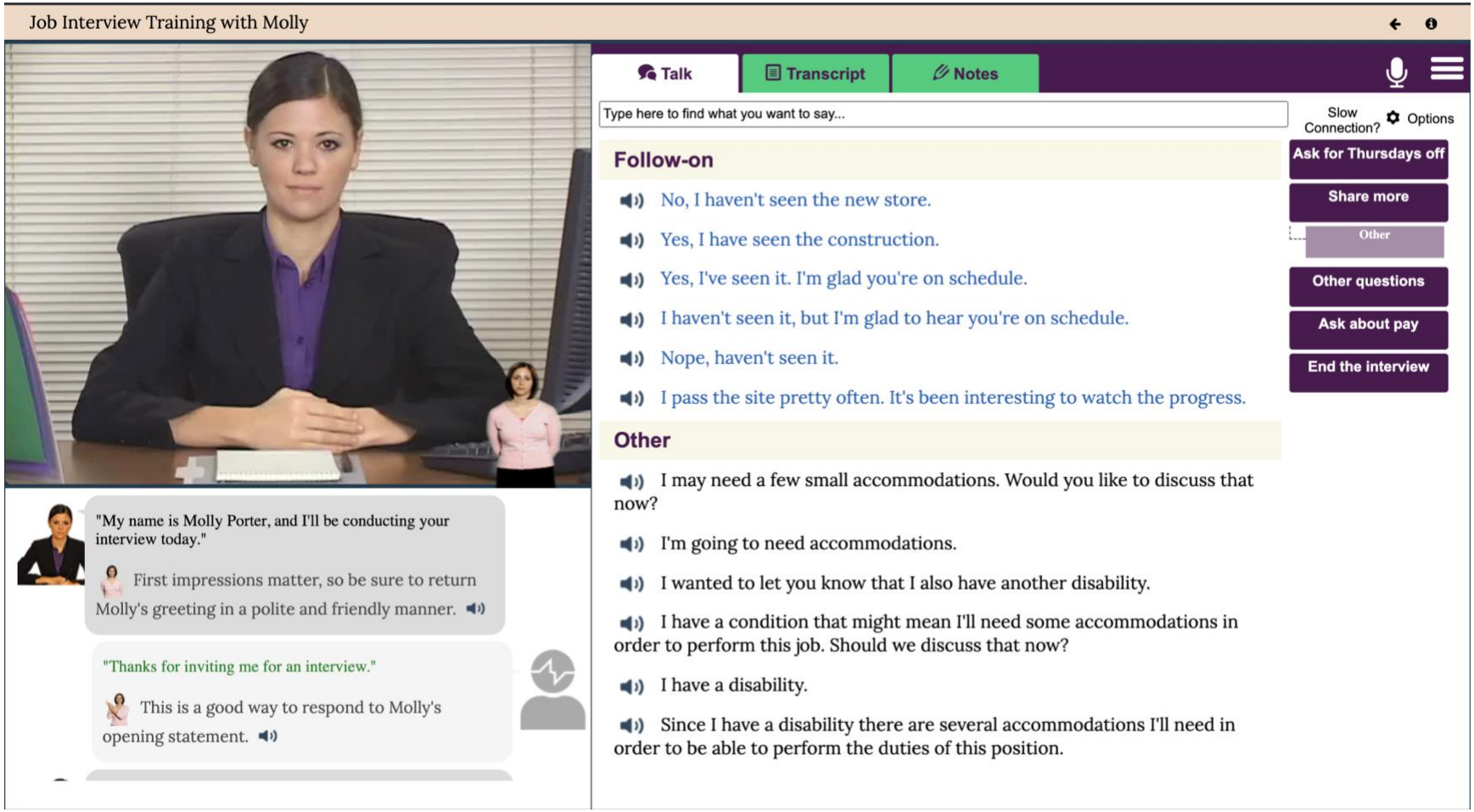
7. Smith MJ, Smith JD, Fleming MF, et al.: Enhancing individual placement and support (IPS) - Supported employment: A Type 1 hybrid design randomized controlled trial to evaluate virtual reality job interview training among adults with severe mental illness. Contemp Clin Trials 77:86-97, 2019

Supplemental Figure 1. Study CONSORT Diagram



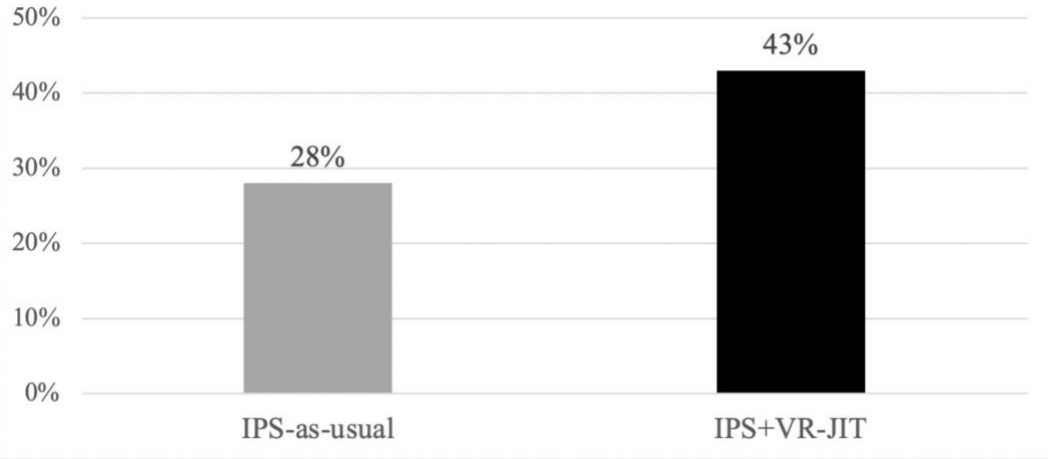
Supplemental Figure 2.

VR-JIT Interface Featuring "Molly Porter"



Supplemental Figure 3

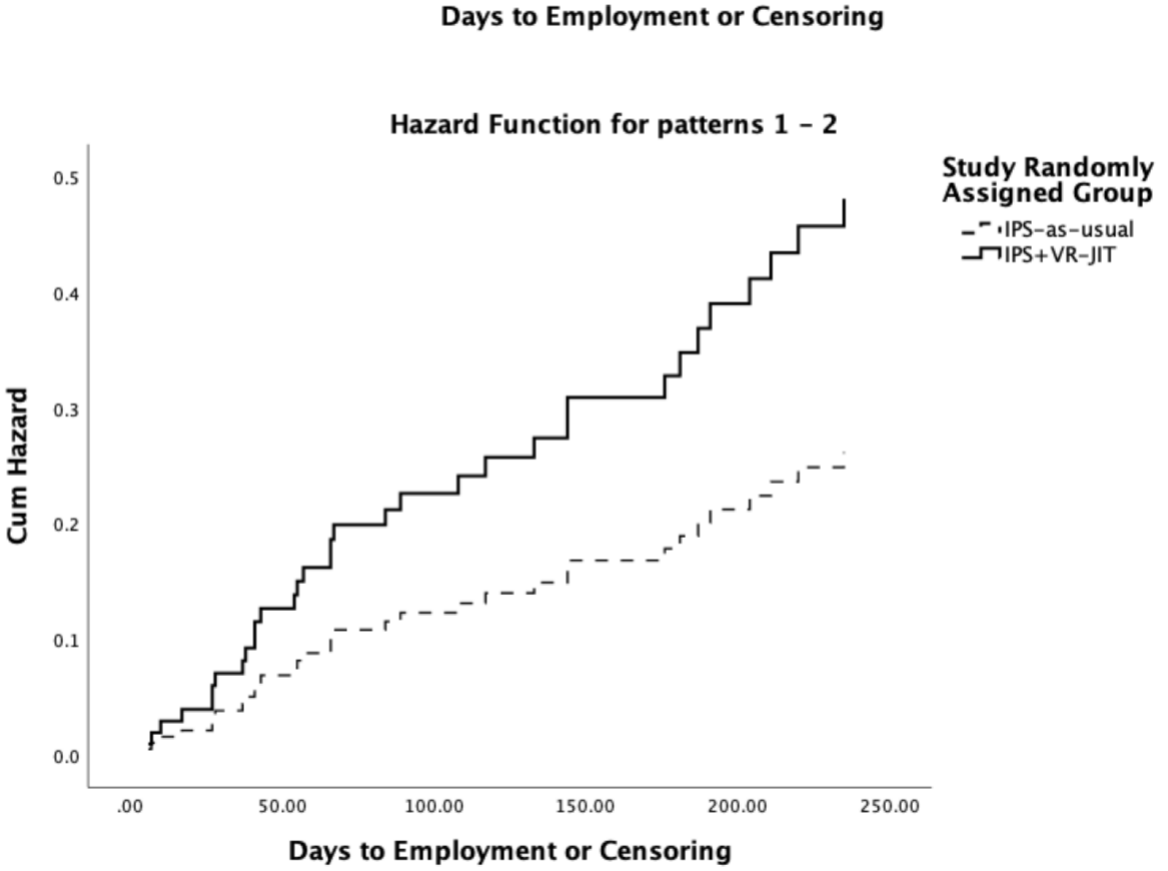
Results of Intent-to-Treat Chi-Square Analysis Evaluating Employment (N=90)[a]



[a] χ^2=2.04 (1), *p*=.076

Supplemental Figure 4

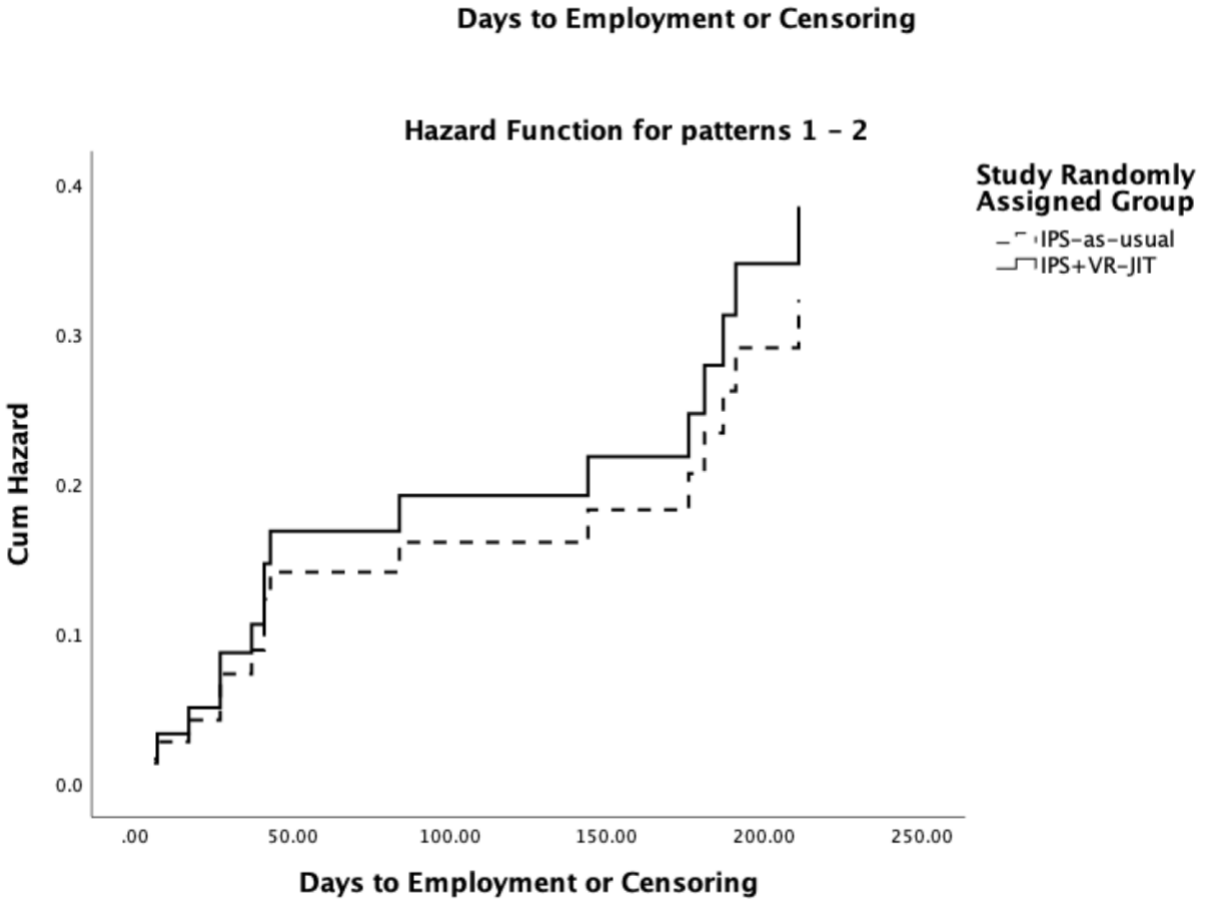
Cumulative Hazard of Employment Across Time[a] By Intent-to-Treat Study Condition (N=90)



[a]Days from randomization to employment. Data were right-censored after 9 months (274 days), but no new employment occurred after day 235.

Supplemental Figure 5

Cumulative Hazard of Employment Across Time[a] By Study Group Among Recent IPS Enrollees (N=44)



[a]Days from randomization to employment. Data were right-censored after 9 months (274 days), but no new employment occurred after day 235.

Supplemental Tables

Supplemental Table 1

Primary outcomes: odds of obtaining employment by nine-month follow-up via multivariable logistic regression for IPS nonresponders (n=46)

Predictor variables	*OR*	95% CI	*p*[b]
Group (reference: IPS as usual)	5.82	1.56[a]	.014
Baseline year	1.64	0.64-4.22	.305
Problematic substance use (reference: no)	3.90	0.80-19.03	.093
Social cognition	1.05	0.99-1.11	.108
Community functioning	0.97	0.91-1.02	.235
Negative symptoms (anergia)	0.72	0.41-1.25	.240

Primary outcomes: odds of obtaining employment by nine-month follow-up via multivariable logistic regression for recent IPS enrollees (n=44)

Predictor variables	*OR*	95% CI	*p*[b]
Group (reference: IPS as usual)	1.05	0.24[a]	.477
Baseline year	4.64	1.13-19.01	.033
Problematic substance use (reference: no)	1.53	0.15-16.15	.723
Social cognition	1.06	0.99-1.13	.074
Community functioning	0.91	0.84-0.99	.021
Negative symptoms (anergia)	0.77	0.53-1.13	.184

Note. *OR* = odds ratio; CI = confidence interval. IPS nonresponder logistic regression model fit statistics were χ^2 (6) = 15.65, p =.016; Nagelkerke R^2 = 0.394. Recent IPS enrollees logistic regression model fit statistics were χ^2 (6) = 18.53, p =.005; Nagelkerke R^2 = 0.470.

[a]CI for directional intervention hypothesis only uses a lower limit confidence interval. [b]1-sided p-value for directional intervention hypothesis, 2-sided p-value for covariates/factors.

Supplemental Table 2

Primary outcome: time-to-employment by nine-month follow-up via multivariable Cox proportional hazards model for IPS nonresponders (n=46)

Predictor variables	*HR*	95% CI	*p*[b]
Group (reference: IPS as usual)	2.70	1.03[a]	.044
Baseline year	2.16	1.08-4.31	.332
Problematic substance use (reference: no)	3.55	1.05-12.04	.047
Social cognition	1.05	1.01-1.09	.087
Community functioning	0.95	0.91-0.99	.136
Negative symptoms (anergia)	0.78	0.57-1.04	.328

Primary outcome: time-to-employment by nine-month follow-up via multivariable Cox proportional hazards model for recent IPS enrollees (n=44)

Predictor variables	*HR*	95% CI	*p*[b]
Group (reference: IPS as usual)	1.19	0.44[a]	.466
Baseline year	1.83	0.90-3.72	.093
Problematic substance use (reference: no)	1.32	0.32-5.49	.700
Social cognition	1.04	1.00-1.07	.050
Community functioning	0.97	0.92-1.01	.152
Negative symptoms (anergia)	0.86	0.64-1.17	.349

Note. *OR* = odds ratio; *HR* = hazard ratio; CI = confidence interval. IPS nonresponder Cox proportional hazards model overall fit statistic was χ^2 (6) = 15.13, p = .019.

Recent IPS enrollees Cox proportional hazards model overall fit statistic was χ^2 (6) = 14.16, p = .028.

[a]CI for directional intervention hypothesis only uses a lower limit confidence interval. [b]1-sided p-value for directional intervention hypothesis, 2-sided p-value for covariates/factors.

Supplemental Table 3

Pearson correlations among Variables (included or considered) for the Intent to Treat Analysis of Employment Outcomes (n=90)

	1	2	3	4	5	6	7	8	9
1. Employment by 9-month follow-up	--	--	--	--	--	--	--	--	--
2. Group status (1=IPS+VR-JIT, 0=IPS-as-usual)	.15	--	--	--	--	--	--	--	--
3. Baseline year	.23*	.03	--	--	--	--	--	--	--
4. Problematic substance use	.15	-.11	-.12	--	--	--	--	--	--
5. Social cognition	.30**	.05	.22	-.11	--	--	--	--	--
6. Community functioning	-.15	.11	.16	-.17	.28**	--	--	--	--
7. Negative symptoms (anergia)	-.18+	.01	.07	.05	-.13	-.10	--	--	--
8. Prior employment (1=employed, 0=not employed, within past 2 years)	.16	-.14	.05	.28**	.14	.03	-.19+	--	--
9. Neurocognition	.05	.02	.07	.02	.23*	.10	-.08	.17	--

+p<.10, *p<.05, **p<.01

Supplemental Table 4

Pearson correlations among variables (included or considered) for the Analysis of Employment Outcomes in IPS nonresponders (top of table; n=46) and recent IPS enrollees (bottom of table; n=44)

	1	2	3	4	5	6	7	8	9
1. Employment by 9-month follow-up	--	.34*	.10	.22	.29*	-.07	-.18	.22	.06
2. Group status (1=IPS+VR-JIT, 0=IPS-as-usual)	-.05	--	.05	-.01	.15	.07	.04	-.01	.17
3. Baseline year	.38**	-.03	--	-.12	.22	.25+	.20	-.01	-.07
4. Problematic substance use	.06	-.21	-.09	--	-.03	-.07	-.09	.36*	.18
5. Social cognition	.31*	-.07	.21	-.19	--	.33*	-.10	.27+	.18
6. Community functioning	-.26+	.14	.02	-.31*	.20	--	.03	.11	.09
7. Negative symptoms (anergia)	-.20	-.03	-.05	.07	-.17	-.23	--	-.04	-.20
8. Prior employment (1=employed, 0=not employed, within past 2 years)	.09	-.23	.19	.10	.06	-.03	-.27+	--	.22
9. Neurocognition	.05	-.16	.22	-.15	.28+	.11	-.03	.17	--

+p<.10, *p<.05, **p<.01