IGNITE, Statistical Analysis Plan

Version 1: February 6, 2026

Clinical Trials.gov registration: https://clinicaltrials.gov/study/NCT05541653

I. Outcome Measures

The primary and secondary outcomes for this study are described in detail in Table 1. Primary outcomes will be drawn from our two-wave survey of study participants (one baseline wave and one endline wave); our secondary outcomes will be drawn from the same survey as well as from key administrative data sources (detailed in Table 1).

Our two primary outcomes will focus on (1) overall health and (2) psychological distress. Our measure of overall health comes from an index based on responses to three survey questions: a self-reported rating of the participant's general health (on a 5 point Likert scale, ranging from poor to excellent); the self-reported number of days in the last 30 in which poor physical or mental health precluded engagement in the participant's usual activities; and a self-reported rating of change in overall health outcomes over the last 12 months (three point scale denoting worse, same, or better). We will use the method of Anderson (2008) to create a single (standardized) index of overall health.[1] This method uses the inverse covariance matrix of the variables to create the index and imputes missing observations at the mean.

Our measure of psychological distress is the Kessler-6 scale, a validated measure of non-specific psychological distress represented by the summed score of six questions querying the frequency of symptoms of feeling nervous, hopeless, depressed, restless, that everything was an effort, and worthless over the past 30 days.[2] The Kessler-6 ranges from 0-24.

II. Power Calculations

For this cluster randomized trial, we assume a Type I error rate of 0.05 and a desired power of 0.8 for the primary outcomes. Our study will include 60 microclusters (30 intervention, 30 control) with at least 8 participants in each cluster (n=480 across all clusters). We will have no fewer than 480 and can enroll up to 720. The number of participants enrolled will depend on our dynamic enrollment strategy that we've discussed already. We assume an intraclass correlation (ICC) of 0.05. For our primary outcomes (and all survey-based outcomes), we will fit a marginal model using generalized estimating equations (GEE) to estimate the effect of the intervention relative to control. Under these assumptions, we will be able to detect a standardized effect of 0.33 at 80% power.[9] This is likely a conservative estimate of power since we will adjust for the baseline value of the outcome and other covariates associated with the outcome. If we are only able to sample from 50 microclusters (25 per arm; n=400 across all clusters), we will be able to detect a moderate standardized effect size of 0.37 at 80% power. Assuming up to 25% attrition between baseline and endline, we would still be able to detect a standardized effect of 0.36 with 60 total microclusters (30 per arm; n=480 across all clusters at first wave).

III. Statistical Analysis

Individual-level primary and secondary outcomes

For all primary and secondary outcomes drawn from our two-wave survey of up to 720 trial participants, we will assess differences in outcomes between the intervention and comparison groups using Generalized Estimating Equation models (GEE) that account for clustering at the study microcluster level. We will use canonical link functions for all outcomes. All models will adjust for all microcluster-level characteristics used in the covariate constrained randomization (see randomization protocol),[3] any individual characteristics found to be imbalanced between the intervention and control groups at baseline (per a standardized mean difference ≥0.1; list of candidate covariates given below), and the baseline value of the outcome (as long as the baseline variable does not exceed 10% missing).

Individual candidate baseline covariates for SMDs:

1. Race/eth (Non-Hispanic Black; Hispanic; Other)
2. Age (in years)
3. Gender (Male; Female; Other)
4. Marital Status (Married; Single; Sep/Div/Widow)
5. Homeownership (Yes; No)
6. Education (HS or lower; Some College; Bachelors Higher)
7. Employed (Employ/Self Employed; Other)

We will conduct an intention-to-treat analysis: participants that move out of (or across) study clusters, separately engage with community partner services while in the control group, or do not engage with these services while in the treatment group will all be considered exposed to their original treatment assignment. The main estimates reported will be based on complete case analysis.

Given that our study includes two primary outcomes, we will account for multiple comparisons using a stacked GEE approach.[4] This approach adequately controls family-wise Type I errors in a manner that is more efficient than overly conservative approaches such as Bonferroni-Holm.

For the survey-based secondary outcomes, we will report results from separate GEE models, including both unadjusted p-values as well as p-values adjusted for multiple inference (the latter based on the Benjamini-Hochberg method).[5]

We will conduct a variety of sensitivity analyses. First, we will assess for non-random attrition by regressing a binary indicator for each participant equal to 1 if the participant did not complete a follow-up survey on the treatment indicator, microcluster level attributes used in the covariate constrained randomization procedure, as well as key baseline covariates (beginning with the list of candidate covariates listed above), adjusting standard errors for clustering at the microcluster level. In the event of non-random attrition (denoted by a statistically significant coefficient on any covariate in the model), we will use inverse probability weighing approaches (IPW) to address potential biases.

Second, we will estimate models that additionally include binary indicators denoting any CC who conducted interviews and fostered linkages with financial partners for each participant. Third, we will estimate a permutation inference model in which we will run our GEE models with covariates in 1,000 permutations, each of which represents a model in which treatment assignment is randomly assigned (or shuffled) 1:1 at the microcluster level (95% CIs for this method, which does not assume a specific error structure, will be constructed a the interval of estimates ranging from the 2.55th to the 97.5th percentile of the distribution of 1,000 point estimates).[6] We will repeat the permutation inference model without covariates and compare estimates from our main models to estimates from these nonparametric approaches. Differences would suggest potential unobserved confounding or model misspecification. In the event of substantive differences, we will estimate bounds on the ITT effects.[7] Fourth, to put bounds on the impact of the intervention, we will impute best case and worst case values for the primary outcome variables with missing values (before constructing the primary index).

Given that enrollment in some clusters occurred contemporaneously with or after the environmental interventions were initiated due to seasonal constraints, one concern is whether exposure to the environmental interventions influenced the types of participants who enrolled in the intervention microclusters. To assess this descriptively, we will report baseline characteristics by arm and enrollment group, as well as calculate SMDs for those characteristics among each of the two enrollment groups. In addition, we will do two sensitivity analyses: first we will run the main model but exclude intervention clusters where enrollment happened after greening. Second, we will include cluster-level indicators for calendar quarter of enrollment in the model, along with an interaction between calendar quarter and arm.

Finally, because a small number of participants were recruited via referrals from current participants, we will apply respondent driven sampling weights[8] and assess the impact of this adjustment on outcomes.

Microcluster level secondary outcomes

For secondary outcomes based on microcluster-level data – namely the number of crimes and nuisance calls per cluster – we will use a difference-in-differences strategy, which leverages additional power from repeat (quarterly) observations that are available in administrative data.[9] Specifically, we will estimate the two-stage difference-in-differences model by Gardner et al (2024) for a secondary outcome, adjusting for calendar time (quarter-year) and cluster fixed effects.[10] The Gardner et al (2024) estimator accounts for staggered adoption, i.e., units entering treatment at different times. We will estimate both an overall average treatment effect on the treated (ATT) and dynamic treatment effects, using an event study specification.[11] The time period of analysis will comprise the 8 quarters prior to intervention and all quarters over the course of the intervention period, which comprises the duration between the end of enrollment and baseline survey data collection to the beginning of follow-up data collection, and the post-intervention period, which all quarters between the beginning and end of follow-up survey data collection. (As data becomes available, we will, in the future, also conduct separate analysis that includes periods 8 quarters after the end of survey data collection, once these data become available. This analysis will be viewed as an extension of our primary difference-in-differences model to assess whether treatment effects were sustained after intervention period ended.)

The coefficients in the event study model that denote time periods prior to the intervention serve as visual and statistical checks of the parallel trends assumption, which we expect to hold given the randomized selection of intervention and control groups. However, if presented with evidence of potential failures of the parallel trends assumption, we will conduct a specification test in which we will estimate and trend out of the post-intervention period any differential pre-intervention trends in outcomes between the two groups.[11]

For all models, we will adjust standard errors for clustering at the study microcluster level.

Sub-group analyses

For our primary outcomes and survey-derived secondary outcomes, we will estimate the above GEE models separately for the following subgroups:

1. Gender (persons identifying as men vs. persons identifying as women; if the sample size permits, we will also consider separate models for trans- and non-binary individuals)
2. Median socioeconomic status (SES) (create an index of SES based on education, home ownership, household income, and financial well-being at baseline)
3. Age (≥50 versus under 50)

We will assess statistical differences in treatment effects between categories within each subgroup by estimating versions of our main GEE models with an interaction term between the treatment assignment variable and assessing the 95% CI for the interaction between treatment and the subgroup indicator.[12]

We also plan to explore heterogeneity in treatment effects using modern machine learning methods.

Evaluating Mechanisms

In addition to evaluating treatment effects on secondary outcomes, we will assess potential mechanisms underlying intervention effects on the primary outcomes using three *descriptive* (and exploratory) approaches.

First, we will use our survey data to examine the extent to which the non-health secondary outcomes (e.g., food security, financial security, social cohesion, stress, experiences of racism, exposure to crime, engagement with greenspaces, health care utilization) mediate the relationship between the intervention and any impacts on the primary outcomes. Specifically, we will conduct a causal mediation analysis to decompose intervention effects by potential mediators.[13]

Second, we will assess changes in key process measures, namely aggregate information on participation in specific social programs, tax returns and refunds, and credit scores. Per agreement with the community financial organizations partnering on this study, this information would only be available at the aggregate level for the treatment group. However, the degree of engagement in different financial services can help assess which services may have been more important in driving intervention effects. For example, if participants in the treatment group on average received large tax refunds, but did not sign up for new benefit programs, that would implicate the former as a potential mechanism (though without information on counterfactual changes in the outcome in the control group, this cannot be proven).

Third, we will assess whether estimated intervention effects were larger among individuals who were most likely to benefit from different program components. For example, individuals who at baseline were not accessing major public benefits or who had not filed tax returns prior would be more likely to benefit from financial interventions than those already engaged with these activities. Similarly, individuals living in study clusters with relatively larger numbers of abandoned lots, fewer trees, or more abandoned homes would stand to benefit more from the environmental interventions than those who did not. Formally, we will estimate versions of our main GEE models for the primary outcomes in which we will include interactions between the treatment indicator and binary indicators for individuals participating in fewer than sample median number of public benefit programs; not filing (or had someone file on their behalf) a tax return in the previous year; being below the median of the household financial security scale; living in a microcluster below the median in terms of tree canopy; living in a microcluster that is above the median in terms of the number of abandoned lots; and living in a microcluster that is above the median in terms of the number of houses. Positive and statistically significant coefficients on these interaction terms would provide suggestive evidence that the specific intervention that would address the pre-intervention attribute in question helped drive treatment effects. For example, a positive and significant coefficient on the interaction between treatment and living in a cluster with above the median numbers of abandoned lots would suggest that abandoned lot remediation played a role in driving overall treatment effects. Interpretation of these coefficients requires knowledge of actual exposure to treatments; e.g., it would be more credible to conclude that provision of tax preparation services helped drive health outcomes if large numbers of intervention group participants reported using these services.

Table 1. Primary and Secondary Outcome Measures

Primary Outcomes				
Outcome	**Measure**	**Data Source**	**Type of Data**	**Timeframe**

Overall Health Index	Composite index using method of Anderson (2008) based on three questions: rating of overall health (5-pt Likert ranging from poor to excellent); rating of how health has changed in last 6 months (better, same, worse); and number of days in the last 30 where physical or mental health precluded engagement in usual activities (self-care, work, recreation); (Oregon HIE)	Survey	Continuous (index)	Baseline, 24 months
Psychological distress	Kessler-6	Survey	Continuous (scale)	Baseline, 24 months
		Secondary Outcomes		
Outcome (by domain)	**Measure**	**Data Source**	**Type of Data**	**Timeframe**
Health				
Overall health	Rating of overall health (5-pt Likert ranging from poor to excellent) (Oregon HIE)	Survey	Ordinal (poor, fair, good, very good, excellent)	Baseline, 24 months
Poor health	Whether individual reported either poor or fair health to overall health question (Oregon HIE)	Survey	Binary (1 = poor or fair health)	Baseline, 24 months
Change in overall health	Rating of how health has changed in last 6 months (better, same, worse)	Survey	Ordinal (better, same, worse)	Baseline, 24 months
Healthy days	number of days in the last 30 where physical or mental health precluded engagement in usual activities (self-care, work, recreation)	Survey	Continuous (number of days)	Baseline, 24 months
Sleep duration	Number of hours of usual sleep (BRFSS)	Survey	Continuous (number of hours)	Baseline, 24 months
Short sleep	Less than seven hours of usual night sleep (BRFSS)	Survey	Binary (1 = short sleep)	
Health care access and utilization				
Healthcare access	Received all needed care in the last 6 months (BRFSS)	Survey	Binary (1 = received all needed care)	Baseline, 24 months
Finances and Benefit Program Participation				

Financial well-being	Consumer Financial Protection Bureau, Abbreviated Financial Well-being Survey	Survey	Continuous (scale)	Baseline, 24 months
Food insecurity	Current Population Survey Food Security Supplement Screener	Survey	Continuous (scale)	Baseline, 24 months
Income tax filing	Whether or not individual (or someone in household on behalf of individual) filed previous years income tax (yes, planning to file late, no) (internally developed)	Survey	Binary (1 = yes)	Baseline, 24 months
Home ownership	Whether or not individual owns house, condo, or mobile home (Add Health)	Survey	Binary (1 = yes)	Baseline, 24 months
Owing on mortgage	Whether or not individual has remaining mortgage payments (internally developed)	Survey	Binary (1 = yes)	Baseline, 24 months
Total debt	Amount of debt added altogether, not including mortgage. (Add Health)	Survey	Continuous (scale)	Baseline, 24 months
Participation in public medical benefit programs	Participation of a household member (including respondent) in Medicaid, Medicare, Medicare savings, LIS, CHIP, Qualified Health Plans, SelectPlan, other, or none (internally developed)	Survey	Binary indicators for participating in any program (=1) and separate indicators for participating in each program (=1)	Baseline, 24 months
Participation in public food benefit programs	Participation of a household member (including respondent) in SNAP, WIC, Senior Food Box, other, or none (internally developed)	Survey	Binary indicators for participating in any program (=1) and separate indicators for participating in each program (=1)	Baseline, 24 months
Participation in public income support or cash benefit programs	Participation of a household member (including respondent) in TANF, LIHEAP, SSI/SSDI, UI, PA General Assistance, PA Emergency Rental Assistance, EITC, CTC Refugee Cash Assistance, CCIS, PA Child Care Tax Credit, other, or none (internally developed)	Survey	Binary indicators for participating in any program (=1) and separate indicators for participating in each program (=1)	Baseline, 24 months

Participation in public home ownership benefit programs	Participation of a household member (including respondent) in PTRR, Homestead Exemption, LOOP, Basic Systems Repair Program, PA Homeowner Assistance, Phily First Home Program, Philadelphia Home Repair Assistance, other or none (internally developed)	Survey	Binary indicators for participating in any program (=1) and separate indicators for participating in each program (=1)	Baseline, 24 months
Greenspaces and Trees				
Frequency of greenspace engagement	Frequency with which individual visits a greenspace (such as a park, garden, greened vacant lot, trail, or any other outdoor space with vegetation) (adapted from Evenson et al 2013 Environment and Behavior)	Survey	Ordinal (never, rarely, once a month, few times a month, once a week, few times a week, every day)	Baseline, 24 months
Time spent in greenspace	Time spent in a greenspace on a typical day (adapted from Evenson et al 2013 Environment and Behavior))	Survey	Ordinal (30 min or less, 31-60 min, 1-2 hrs, 2+ hrs)	Baseline, 24 months
Reasons for not spending time in greenspace	Things that stop an individual from spending time in greenspace (adapted from Evenson et al 2013 Environment and Behavior)	Survey	Categorical (weather (too cold or too hot), safety concerns, no time, too tired, I don't like spending time outside, other, nothing stops me from spending time outside in greenspace)	Baseline, 24 months
Perception of tree cover	Beliefs about number of trees in the neighborhood (internally developed)	Survey	Categorical (need more trees, enough trees, need less trees, unsure)	Baseline, 24 months
Tree planting concerns	Whether or not individual has concerns about planting more trees in neighborhood (internally developed)	Survey	Binary indicators for any concerns (=1)	Baseline, 24 months

Perceived tree health benefits	Whether or not individual believes trees confer health benefits (e.g., safety, mental health benefits, physical health benefits, social benefits, environmental benefits, aesthetic benefits)	Survey	Binary indicators denoting belief of any health benefits (=1) and separate indicators for each type of benefit	Baseline, 24 months
Stress and Agency				
Perceived stress	Perceived Stress Scale	Survey	Continuous (scale)	Baseline, 24 months
Neighborhood Perceptions				
Time spent in neighborhood	If individual endorses spending time relaxing, socializing, or hanging out in porches, stoops, and front yards of neighborhoods (adapted from Kahneman et al 2004 Science)	Survey	Categorical (5-pt Likert, Strongly agree to Strongly disagree)	Baseline, 24 months
Neighborhood social capital	Neighborhood Social Cohesion & Exchange and Social & Physical Disorder Scales	Survey	Continuous (scales)	Baseline, 24 months
Physical disorder	Whether or not participant reports a lot of abandoned buildings in their neighborhood (Ross and Mirowksi)	Survey	Binary (1 = yes)	Baseline, 24 months
Microcluster-Level Outcomes				
Neighborhood crime rates	Overall number of crimes, number of violent crimes, serious crimes, and gun related incidents	Philadelphia Police Dept. Crime Data - open access	Continuous (rate)	Quarterly data from 8 quarters prior to enrollment and 4 quarters after intervention period complete
Nuisance calls	Number of 311 calls and number of 311 calls for neighborhood cleanup and remediation-related issues	City of Philadelphia - open access data	Continuous (rate)	Quarterly data from 8 quarters prior to enrollment and 4 quarters after intervention period complete

References

1. Anderson ML. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *J Am Stat Assoc*. 2008;103(484):1481-1495. doi:10.1198/016214508000000841

2. Kessler RC, Green JG, Gruber MJ, et al. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *Int J Methods Psychiatr Res*. 2010;19(S1):4-22. doi:10.1002/mpr.310

3. An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes - Li - 2017 - Statistics in Medicine - Wiley Online Library. Accessed January 21, 2026. https://onlinelibrary.wiley.com/doi/full/10.1002/sim.7410

4. Simultaneous inference for multiple marginal generalized estimating equation models - Robin Ristl, Ludwig Hothorn, Christian Ritz, Martin Posch, 2020. Accessed January 21, 2026. https://journals.sagepub.com/doi/full/10.1177/0962280219873005

5. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

6. Braga AA, MacDonald JM, McCabe J. Body-worn cameras, lawful police stops, and NYPD officer compliance: A cluster randomized controlled trial*. *Criminology*. 2022;60(1):124-158. doi:10.1111/1745-9125.12293

7. Rosenbaum P. *Observational Studies*. 2nd ed. Springer; 2002.

8. Schonlau M, Liebau E. Respondent-Driven Sampling. *Stata J*. 2012;12(1):72-93. doi:10.1177/1536867X1201200106

9. McKenzie D. Beyond baseline and follow-up: The case for more T in experiments. *J Dev Econ*. 2012;99(2):210-221. doi:10.1016/j.jdeveco.2012.01.002

10. Gardner J, Thakral N, Tô LT, Yap L. Two-Stage Differences in Differences. Published online May 2025.

11. Goodman-Bacon A. Difference-in-differences with variation in treatment timing. *J Econom*. 2021;225(2):254-277. doi:10.1016/j.jeconom.2021.03.014

12. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *N Engl J Med*. 2007;357(21):2189-2194. doi:10.1056/NEJMsr077003

13. Imai K, Keele L, Yamamoto T. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Stat Sci*. 2010;25(1):51-71. doi:10.1214/10-STS321