Best Practices to Prevent COVID-19 Illness in Staff and People with Serious Mental Illness and Developmental Disabilities in Congregate Living Settings

NCT04726371

Statistical Analysis Plan

Version date: March 22, 2023

Co-Primary Outcome: Effectiveness (COVID-19 infection)

Outcomes

New COVID-19 group home incidence (new laboratory-confirmed COVID-19 cases among residents, staff (together and then separately). Home-level COVID-19 infection incidence rates per 100 person-months were calculated by home and time period using the infection counts and follow-up time.

Covariates

- *Study arm:* General Best Practice or Tailored Best Practice
- *Agency:* Unique identifier of the six provider organizations.
- *Group Home*: Unique identifier of the group home where resident lived or staff worked.
- *Stratification factors*: Stratification factors used for randomization, derived from the baseline racial and ethnic composition of group home residents and staff and a COVID-19 infection risk score.
- *Follow-up time (exposure):* Follow-up time in months for residents in a home was calculated using dates of residence based on intake and discharge data. Follow-up time for staff in a home was calculated using dates worked based on payroll data. Individual staff could contribute time to multiple homes during the same time period. Follow-up time for each home and follow-up period was calculated by totaling the resident and/or staff follow-up time in the home and period.
- *COVID-19 infection case count:* Resident and staff infection counts were determined to calculate incidence rate at the level of group home by time period. Multiple infections per-participant were counted if cases were at least 3 months apart. Resident infections were counted towards the home where they lived at time of infection. Staff often work in multiple homes concurrently. Since staff are being exposed to and/or exposing multiple home environments to infection, their case counts were allocated to any homes where they worked during the follow-up period they were infected.
- *Baseline incidence rate:* Home-level COVID-19 infection incidence rates per 100 person-

months were calculated by home and time period using the infection counts and follow-up time at baseline.

Sample Size Calculations and Power

For power calculations we assumed a 2-tailed test with significance level .05, intraclass correlation (ICC) of 0.02 within the four randomization clusters (i.e., highest and lowest prevalence GHs within GBP and TBP), and ICC of 0.50 for repeated measures at the GH-level. Given COVID-19 incidence as of May 2020 rates based on available COVID-19 prevalence data for GHs residents and staff, we will have at least 80% power to detect 8.7 and 6.3% percentage point difference of COVID-19 infection between GBP and TBP for residents (ID/DD combined) and staff, respectively. We used conservative assumptions, given the possibility of undetected COVID-19 in GHs. The primary outcome is COVID-infections in both staff and residents with correction for type I error applying Bonferroni's method.

Table 1. Power calculations.

	Estimated COVID-19 Prevalence		Difference	Cohen's d[a]	Power
Population	**Control (GBP)**	**Intervention (TBP)**			
Group Home Staff	13.10%	6.80%	6.30%	0.38	80%
Residents with SMI & ID/DD combined	26.50%	17.80%	8.70%	0.28	80%
Residents with ID/DD	40%	30%	10%	0.24	80%
Residents with SMI	13.10%	6.80%	6.30%	0.38	80%

[a]Small effect=0.2; Medium effect=0.5

Time Frame

These analyses include resident and staff data from January 1, 2021 - July 31, 2021. Data was collected across six time points:

- Baseline: 1/1/2021 – 3/31/2021
- 3-month follow-up: 4/1/2021 – 6/30/2021
- 6-month follow-up: 7/1/2021 – 9/30/2021

- 9-month follow-up: 10/1/2021 – 12/31/2021
- 12-month follow-up: 1/1/2022 – 3/31/2022
- 15-month follow-up: 4/1/2022 – 7/30/2022

Data Collection and Sources

Person-level (individual health worker and resident) measures were collected at baseline, 3-, 6-, 9-, 12-, and 15-months post-baseline. These limited datasets included information from routinely collected administrative data at each participating agency, such as dates of COVID-19 events for staff and residents (i.e., dates of positive test results, hospitalizations, deaths, and vaccinations). We also received the unique, encrypted identifiers of the group home(s) that residents lived in and staff worked in during each 3-month data collection period. We also obtained information on demographic characteristics (e.g., age, gender, race/ethnicity) for residents and staff. All data were transmitted from the provider organizations to Massachusetts General Hospital using a secure e-mail portal.

Analytical and Statistical Approaches

The raw incidence rates were summarized across the homes by treatment group (Tailored Best Practices TBP vs Generic Best Practices GBP) for each time period. To compare the effectiveness between these groups, we fit a Poisson generalized linear mixed model with log-link for the group home level infection incidence rates over the 5 follow-up periods. The model included main effects for treatment group, linear and quadratic time trends over the follow-up periods, and interaction effects between treatment and the time periods. The model additionally included fixed effects for randomization factors, baseline incidence at each home, and group home agency, along with a group-home level random intercept.

A joint Wald test was performed to assess whether there were significant differences in the trends between treatment groups based on this model for the main effect for treatment including interactions between treatment and time trends. Estimates of the conditional incidence rate ratio (IRR) at each time point were calculated based on the linear combination of

the treatment main effect and treatment by time interaction effects. Correction for type I error was controlled by Bonferroni's method. Marginal mean incidence rates by treatment group per 100 person-months were estimated based on the model assuming 0 random effect. These analyses were repeated for various subgroups: staff, residents, IDD residents, and SMI residents separately. The analyses were performed in Stata 16.

Changes to the Original Study Protocol

We used an ANCOVA model in which outcomes observed at baseline were adjusted for as a covariate rather than included as a follow-up time point because of discrepancies in the definition of the fidelity score between the baseline and follow-up time points (resident masking was removed after baseline due to policy change). This approach was adopted for other endpoints to maintain consistency across all primary analyses. A quadratic term for time was included in the model to improve the model fit in accordance to observed trends.

Co-Primary Outcome: Implementation

Outcomes

Fidelity to COVID-19 prevention practices among residents and staff (masking, screening, handwashing, vaccinations, booster vaccinations).

Covariates

- *Study arm:* General Best Practice (GBP) or Tailored Best Practice (TBP)
- *Agency:* Unique identifier of the six provider organizations.
- *Baseline Fidelity score:* Fidelity scores were calculated by home at baseline.
- *Group Home*: Unique identifier of the group home where resident lived or staff worked.
- *Stratification factors*: Stratification factors used for randomization, derived from the baseline racial and ethnic composition of group home residents and staff and a COVID-19 infection risk score.

Sample Size Calculations and Power

For power calculations we assumed a 2-tailed test with significance level .05, intraclass correlation (ICC) of 0.02 within the four randomization clusters (i.e., highest and lowest prevalence GHs within GBP and TBP), and ICC of 0.50 for repeated measures at the GH-level. For the primary implementation outcome, the continuous Best Practices Fidelity score, we will have 80% power to detect minimum 0.21 standardized mean difference between intervention and control (Cohen's d = 0.21). For a dichotomous fidelity threshold of 80% Best Practices Fidelity, we will have 80% power to detect 7.2% difference (Cohen's d = 0.33).

Time Frame

These analyses included home-level Program Director (PD) responses to 6 waves of Fidelity surveys distributed between May 1, 2021 - October 31, 2022. Measures were collected during the following time periods:

- Baseline: 5/1/2021 – 7/31/2021
- 3-month follow-up: 8/1/2021 – 10/31/2021
- 6-month follow-up: 11/1/2021 – 1/31/2021
- 9-month follow-up: 2/1/2022 – 4/30/2022
- 12-month follow-up: 5/1/2022 – 7/31/2022
- 15-month follow-up: 8/1/2022 – 10/31/2022

Data Collection and Sources

PDs at study sites were completed a brief 10-minute COVID survey covering prevention practices used in the prior 3 weeks. Responses were assessed at baseline, 3-, 6-, 9-, 12-, and 15-months. The survey items were designed in collaboration with care staff and other stakeholders to support the relevance, feasibility, and comprehension of the measures in the routine care context. Surveys were filled out by PDs on a secure web application (REDCap). PDs who completed the survey received $10 gift cards.

Analytical and Statistical Approaches

Fidelity measures: Fidelity measures were determined by state level policies in place while the survey was in the field. Since COVID-19 prevention practice policies changed throughout the study period, our definition of Fidelity is dynamic, shifting across time points. The following measures were counted for each time point:

Table 2. Relevant Fidelity measures by wave, based on current COVID-19 policies.

	Vaccinations		Boosters		Screening		Hand Washing		Masking	
Time	**Residents**	**Staff**	**Residents**	**Staff**	**Residents**	**Staff**	**Residents**	**Staff**	**Residents**	**Staff**
Baseline	☑	☑			☑	☑	☑	☑	☑	☑
3-month follow-up	☑	☑			☑	☑	☑	☑		☑
6-month follow-up	☑	☑			☑	☑	☑	☑		☑
9-month follow-up	☑	☑	☑	☑	☑	☑	☑	☑		☑
12-month follow-up	☑	☑	☑	☑			☑	☑		☑
15-month follow-up	☑	☑	☑	☑			☑	☑		☑

We followed methodology for developing fidelity scales for evidence-based practices described by Bond and colleagues in preparing the TBP and GBP fidelity scales. We operationally defined 2-4 items assessing each measure, with items scored on a 5- to 6-point continuum with a rating of 5 or 6 indicating full adherence to the fidelity standard and 1 indicating complete lack of adherence. Responses were only included in the analysis if all items were completed. If a home contributed 2 PD responses in one time period, then the responses were averaged.

A home's overall, continuous fidelity score was calculated by averaging measure-specific scores by time period. Each active measure was given equal weight. For example, at the 6-month time point, Vaccination, Screening, Handwashing, and Masking (for staff) were active policies. The percentage scores for each of these measures were then averaged together. As such, the overall fidelity score ranged from a low of 0% to a high of 100%.

Statistical analysis: The mean fidelity scores were summarized across the homes by treatment group (TBP vs GBP) for each time period. We fit a linear mixed model for group home-level fidelity scores over the 5 follow-up periods. The model included main effects for treatment group, linear and quadratic time trends over the follow-up periods, and interaction effects between treatment group and time periods. The model additionally included fixed effects for

the randomization factors, baseline fidelity, group home agency, in addition to a group home level random intercept. Homes with no baseline response were excluded from the analysis. For an additional evaluation of fidelity (based on a dichotomous outcome of a threshold of achieving at least 80% fidelity), logit link and binomial distribution will be specified.

A joint Wald test against the null showed that there were no differences in the trends in mean fidelity scores was conducted. Estimates of the differences in mean fidelity score between treatment groups at each time points were calculated based on linear combinations of the treatment main effect and treatment by time interaction effects. Correction for type I error will be controlled by Bonferroni's method. The marginal mean score by treatment group were estimated based on the model.

Changes to the Original Study Protocol

We used an ANCOVA model for the same reasons stated earlier. Quadratic time was added to the model due to observed trends in the data and conditional IRR estimates were generated.