Cover Page

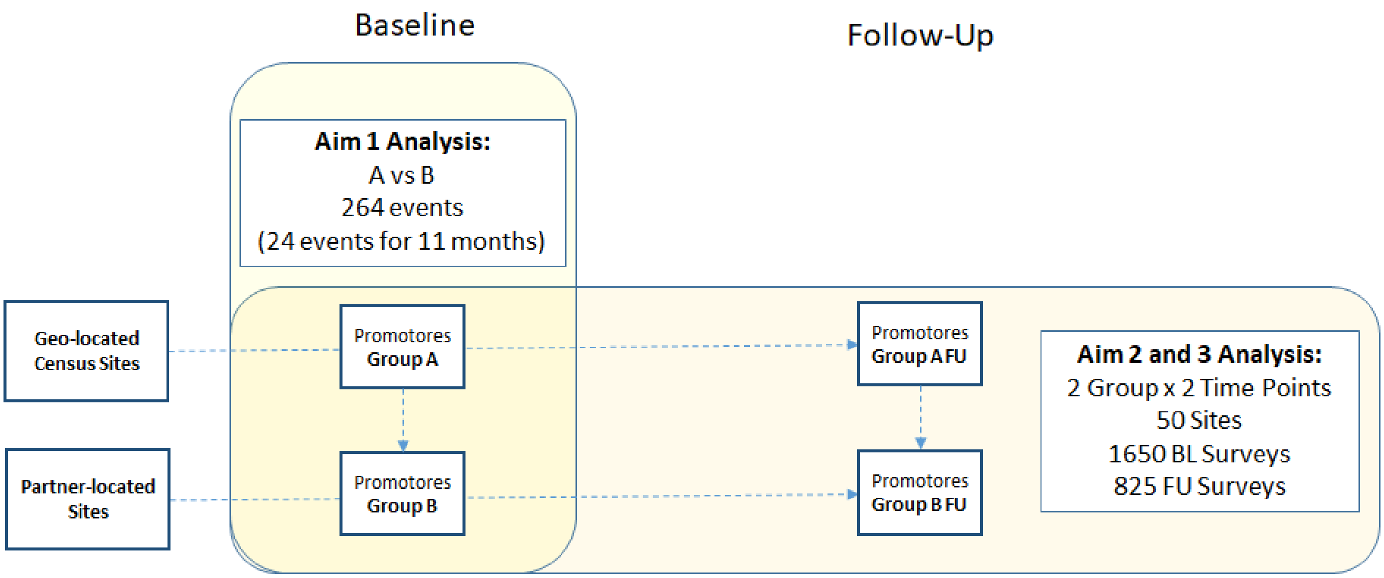
Document:	Statistical Analysis Plan
Document Date:	2021-05-27

Project Title:	COVID-19: Healthy Oregon (Oregon Saludable): Together We Can (Juntos Podemos) Phase II
Identifiers:	NCT05082935
Unique Protocol ID:	10032020.002.2
Secondary ID:	P50DA048756-03S3

4.3 Statistical Design and Power

A power convention of .80 and for an alpha level of .05 was chosen for evaluating the multilevel model for the proposed site level outcomes (Aim 1) and for the individual level outcomes (Aim 2 and Aim 3). The minimally detectable effect size (MDE) was estimated using Monte Carlo simulation with 1000 replications in M*plus* 8.5. An advantage of the Monte Carlo simulation approach is the ability to estimate parameter bias, standard errors, and coverage of confidence intervals.

Figure 4.3.1 Phase II Study Design and Expected Sample Sizes by Unit of Analysis



Aim 1 Analyses

Converting the Odds Ratio from the site level preliminary studies and the per capita outcomes Cohen's *d* ranged from .45 to .60, medium effects. The primary outcome for Aim 1 will be the number of Latinx tested specified as a generalized linear multilevel Poisson model for the count model and as a general linear model for the per capita proportional measures. For Aim 1 analyses, the unit of analysis is the number of total events nested within site. Across the continuing Phase I and the proposed Phase II testing events, there will be a total of 24 events per month for a total of 11 months (264 events).

As an illustration of total number tested, formally, the primary generalized linear model will be the following:

Level 1: $\text{E(Test Count)} = \lambda$

$$\text{Log}[\lambda] = \beta_0 + \beta_1(\text{Historical Week}) + r$$

Level 2:

$$\beta_0 = \gamma_{00} + \gamma_1(\text{Location Contrast}) \ldots \gamma_2(\text{Site Charateristics}) + u_0$$
$$\beta_1 = \gamma_{10} + u_1$$

Mixed Model:

$$\text{Log}[\lambda] = \gamma_{00} + \gamma_{01}(\text{Location Contrast}) + \gamma_{02}(\text{Site Characteristics}) + \gamma_{10}(\text{Historical Week}) + u_0 + u_1 + r$$

Where the expected count of Latinx tested represented as log λ is regressed on level 1 historical calendar week β1 and the effect of the site-location outreach contrast γ1 is estimated at level 2 as location type effect, controlling for socio-demographic census characteristics of each site γ2. Nested models will be compared and evaluated for over or under dispersion, zero-inflation, and normality for the linear mixed model estimating per capita Latinx tested.

Aim 1 Power. To illustrate power, MDE was estimated for proportion of Latinx tested per capita Latinx population with 1000 replications. One advantage of M*plus* is the ability to specify events per month according to the design in which 10 census-located sites from Phase 1 will be conducting 20 events per month; and in the other study arm, 4 events per month will be conducted via community partner-located sites. Additionally, the group contrast for location type is simulated according to the clustering design with the appropriate allocation (mean = .42 and variance = .25). Residual variance of number of tested Latinx was specified at .7 for level 1 and .5 for level 2 in order to estimate an intra-class correlation (ICC) of .40, a conservative parameter reflecting the site level ICC from the preliminary analyses ranging from .24 to .37 for the related Latinx testing outcomes described in the preliminary studies section. Covariate parameters were entered with and effect size (ES) of .30 for historical week, and .19 for site characteristics at level 2 with .91 coverage. Models showed that the study is powered .81 to minimally detect an ES of .50, a medium effect.

Aim 2 and Aim 3 Analyses

Both Aim 2 and Aim 3 analyses will focus on the individual unit of analysis. Aim 2 will evaluate whether Phase II implementation fidelity of the *Promotores de Salud* health intervention will improve testing, reduce vaccination hesitancy, and increase vaccine acceptance and health literacy. Aim 3 will focus on the social determinants of pre-post change in the likelihood of vaccination, repeated tested for those unvaccinated, and reduced vaccine hesitancy. Both Aims will be specified as within-intervention group pre-post change. In general, we will estimate pre-post linear mixed models for vaccine hesitancy and generalized linear mixed models for increased testing counts and vaccine acceptance.

Using vaccination as an outcome, both Aims 2 and 3 are formally represented as a multilevel Bernoulli logistic model predicting the likelihood of vaccination:

Level 1:
$$\text{Prob(Vaccine=1)} = \Phi$$
$$\text{Log}[\Phi/(1-\Phi)] = \beta_0 + \beta_1(\text{BL Vaccine Status}) + \beta_2(\text{Social Determinants}) + \beta_3(\text{Historical Week}) + r$$

Level 2:
$$\beta_0 = \gamma_{00} + \gamma_1(\text{Fidelity of Implementation}) + \gamma_2(\text{Location Contrast}) \ldots \gamma_n(\text{Site Charateristics}) + u_0$$
$$\beta_1 = \gamma_{10} + u_1$$
$$\beta_2 = \gamma_{20} + u_2$$
$$\beta_2 = \gamma_{20} + u_3$$

Mixed Model:
$$\text{Log}[\Phi/(1-\Phi)] = \gamma_{00} + \gamma_{01}(\text{Fidelity}) + \gamma_{02}(\text{Location Contrast}) + \gamma_{02}(\text{Site Characteristic}) + \gamma_{10}(\text{BL Vaccine Status}) + \gamma_{20}(\text{Social Determinants}) + \gamma_{30}(\text{Historical Week}) + u_0 + u_1 + u_2 + u_3 + r$$

Where the probability of getting vaccinated Φ is regressed on level 1 baseline vaccine status β_1, which then estimates the change in vaccine status from pre- to post-*Promotores*. Aim 2 fidelity of implementation is then estimated at level 2 as γ_1 controlling for location type and site level characteristics. Aim 3 social determinants are estimated at level when shown here as β_2.

Aims 2 and 3 Power. For the multilevel model addressing aims 2 and 3, based on our preliminary Phase I data, we expect a total of 1650 survey participants over 11 months (200 tests per month via census-located events and 100 tests per month in partner-located events, for a total of 300 tests/month, with a 50% enrollment rate into the survey × 11 months). For the one-month follow-up we anticipate a 50% retention rate. For a total of 825 pre-post participants. Power was estimated using a two-level Monte Carlo simulation, with the ability to model pattern missingness via both location approaches. To be conservative a retention rate was estimated at .50 for the established Phase I location approaches which will include a higher frequency of events, and a retention of .40 was estimated as the missing data pattern for Phase II partner-located; moreover, the proportion of missing data patterns were specified using the proportional design allocation and an ICC of .30 for individual testing outcomes. The proposed study is powered .81 to detect a MDE of .32 for fidelity of implementation at level 2, and is powered .83 to detect an MDE of .11 for social determinant variables at level 1 with .92 coverage.