STATISTICAL ANALYSIS PLAN

REOpening Schools Safey and Educating Youth (ROSSEY Study)

NCT04859699

Version 2.0
November 26, 2024

1. Introduction to the SAP

This statistical analysis plan (SAP) describes the plans for statistical analyses which will address the protocol objectives of the study, except the qualitative research methods, which are to be addressed using non-statistical qualitative data analytic methods.

2. Study Schema and objectives

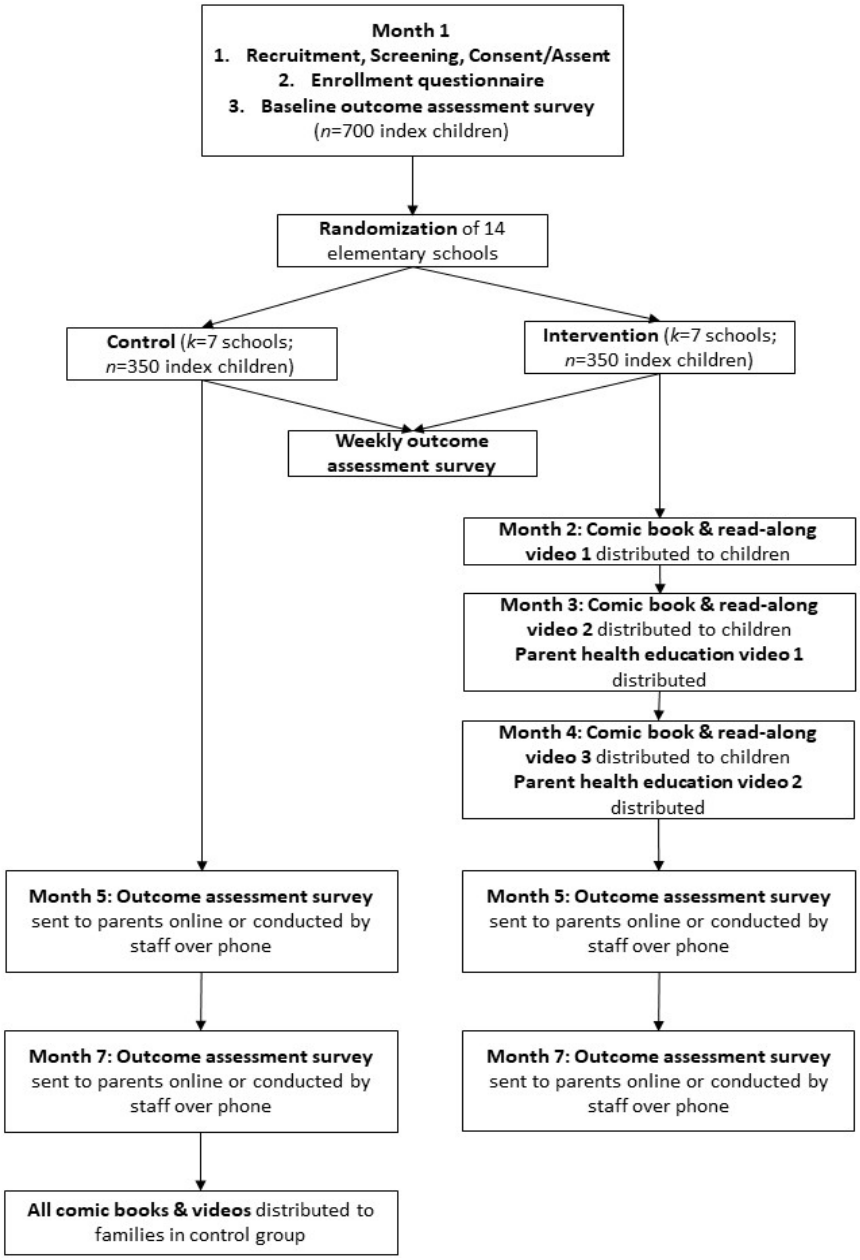
Purpose:	To determine the effectiveness of risk communication for parents and students on students' attendance to school, COVID-19 testing and vaccine uptake
Design:	A clustered randomized controlled trial of risk communication intervention
Central Hypotheses:	Our central hypotheses are that students and parents who view the risk communications will be more likely to have less school absence, increased covid-19 testing and vaccine uptake.
Primary Objectives:	• To determine if implementing three COVID-19 education comic books for children and two COVID-19 education videos for parents will lower school absenteeism, increase COVID-19 testing, and vaccine uptake in children.
Secondary Objectives:	• To determine if implementing three COVID-19 education comics books for children and two COVID-19 education videos for parents will increase physical activity, improve sleep, increase emotional regulation in children. • To determine if implementing three COVID-19 education comics books for children and two COVID-19 education videos for parents will increase physical activity, improve sleep, decrease depression and anxiety in parents
Population:	Children ages 5-13 enrolled in kindergarten to 5th grade in Yakima School District
Study Size:	14 schools randomized; 7 to intervention and 7 to comparison arms with 450 children in each arm; a total of 900 children from 14 schools.
Primary Outcome Measurements:	A. Attendance captured through the school district B. COVID-19 testing data available through the school district testing program and self-reported survey C. COVID-19 vaccine data available through self-reported survey
Study Duration:	2 years including enrollment period, 2 months to deliver the intervention, 5 months for post 1 survey, 7 months weeks for post 2 survey.

3. Study Design

ROSSEY study is a cluster randomized controlled trial to test COVID-19 risk communication intervention with 900 participants from 14 elementary schools in the Yakima School District during

the 2022-2023 school year. Schools are randomized 1:1 (matched by school size and English language learners) to either intervention or control group. Students and parents randomized to the intervention group receive the COVID-19 risk communication intervention. Students and parents randomized to the control group receive the same risk communication intervention at the end of the study as a delayed intervention.

Study Timeline



4. Sample size and power

For month 5 outcome. We will enroll 900 elementary school students from 14 schools in the Yakima School District during the 2022-2023 school year. The sample size of 7 schools per study arm and an average of 50 students per school will provide 90% power to detect a difference of at least 4 hours per week (SD=8.0) of in-person school attendance between the intervention and control groups, while assuming intra-cluster correlation (ICC) of 0.05 and type I error rate of 0.05. The estimated sample size of 14 schools with 900 students provides 80% power to detect a difference of 20% or more in COVID-19 vaccination rates between the two study arms, assuming ICC of 0.05 and type I error rate of 0.05.

Minimum detectable difference in hours per week between intervention arms in proposed cluster RCT assuming enrollment of 15 students per classroom, 80% power and two-sided type I error rate of 0.05.

		Two Arm			
	Number of clusters per arm	30		40	
	ICC	.01	.10	.01	.10
Standard Deviation	5.0	1.0	1.5	.87	1.3
	7.5	1.5	2.2	1.3	1.9

5. Population and Data Source

Participants enrolled in the ROSSEY Study: 900 children and parents from the Yakima School District. Any randomized participants subsequently found to be ineligible will be removed. The data sources for the primary analyses are the attendance records and positive COVID testing results received from the school administration and the survey data as entered in REDCap. COVID Vaccination status data is obtained through the parent and child surveys that are completed in REDCap.

6. Presentation of comparative analyses

For each of the continuous outcomes (including the primary outcome), the mean and standard deviation for each allocated group will be presented, together with the mean between-group difference, 95% confidence interval for the difference and p value. For binary outcomes, the percentage and frequency of children in the outcome category of interest (e.g. percentage vaccinated) will be presented for each allocated group, along with the odds ratio for the intervention effect, 95% confidence interval for the odds ratio and p value.

7. Outcome Definitions

a. Primary Outcomes

The three co-primary outcomes for the trial are school absenteeism, COVID-19 testing, and COVID-19 vaccine uptake. Absenteeism are the student's school attendance data for 2022-2023 school year, COVID-19 data are captured through school testing and self-reported survey, and COVID-19 vaccine uptake data are captured through self-reported survey. All children and parent data after randomization will be included in the intention to treat analysis.

Absenteeism: total number of school days missed by each individual student participant, both for the entire school year and month-by-month, as captured and shared with us by school district administrative data systems. Absenteeism will also be measured by school days missed and will be a count variable.

COVID-19 testing: Health reports from the school are provided for the 2021-2022 and 2022-2023 school year that list positive covid test results of study participants with the date of testing and the school of attendance. Positive COVID test results will be a binary yes/no variable.

COVID-19 vaccine uptake: Parents are asked at enrollment to report their vaccination status and their child's vaccination status. At post survey 1 and post survey 2 the parent is asked to report if they or their child have received any COVID-19 vaccine or booster since enrollment. Vaccination status will be a binary yes/no variable.

b. Secondary Outcomes

Secondary outcomes for the trial include the following outcomes from the PROMIS – Pediatric and Parent Proxy Instruments. The PROMIS instruments used are Children's physical activity (PA), sleep, stress, depression, and anxiety. Parents' PA, sleep, stress, depression and anxiety. The final score is represented by the T-score, a standardized score with a mean of 50 and a standard deviation (SD) of 10.

The outcome measures will be collected at baseline, and months 5 and 7 surveys.

8. Primary exposure definition

In keeping with intention-to-treat analysis principles, the primary exposure of interest is randomization group as determined by the appropriate random allocation list(s). Even if a participant is mistakenly given the incorrect group assignment in implementing the study, they will be analyzed in the group to which they were randomly assigned according to the appropriate random allocation list. Using the random allocation assignment will avoid any actual or appearance of potential selection bias in the comparison groups, but could bias trial results towards the null when such errors are made. All efforts should be made to avoid such errors.

9. Interim analysis plan

For this study there were no interim analysis.

10. Multiple testing

We will adjust for multiple testing by controlling FDR.

11. Missing Data

After enrollment data collection, some missing demographic information was included in post survey one. We used the later data to fill in missing data over multiple time points. We gathered as much information as possible by engaging in multiple follow-up attempts. Incentives were used to have as few participants drop out as possible and to keep participants engaged.

For other missing data at the end, we will inspect missing data mechanisms and the amount of missing data. We will conduct an additional sensitivity analysis with multiple imputations using multivariate imputation with chained equations (MICE) approach. We will report the amount of missing data and apply appropriate methods to account for the missing data.

12. Planned Statistical Analyses

a. CONSORT diagram

Screening and enrollment into the study will be shown in a CONSORT-style diagram, with summary of reasons for ineligibility or non-enrollment; randomization assignment into the two groups, and retention to the 5-month and 7-month outcome assessments.

b. Baseline Demographics

Baseline characteristics of enrolled participants will be summarized by randomization group, and overall, in a table. It is expected that children in both allocated groups will, on average, be similar, given the randomization procedure.

c. Efficacy Analyses Overview

Post survey 1 outcome analyses will be conducted after all participants have been contacted and the survey is closed with missing responses marked as loss to follow-up but before the collection of the second round of surveys began. Post survey 2 outcome analysis will be conducted after all participants have been contacted and the survey is closed.

All efficacy analyses will be performed using the intention-to-treat method in the sense that randomly assigned group will be our exposure regardless of adherence to POC testing offered. In addition, for our combined ART care outcome, we have designed the outcome to minimize missing data for our primary comparisons.

i. Descriptive Analyses:

Positive COVID test results from the school district testing program will be summarized by counts by randomization group. Vaccine status and covid testing results will be summarized by counts and percentages by randomization group. Continuous variables by mean, standard

deviation(SD) and median (IQR) by randomization group. Secondary outcomes will be summarized and compared using Chi-squared tests

ii. Primary Objectives

The primary analysis will be based on the average number of onsite learning hours per week over the study period for each student. The data will be analyzed using a linear mixed effects model with district and randomization arm as fixed effects, and school and classroom within school as random effects. We will first test for any difference between the randomization arms using $\alpha = 0.05$ and report the comparisons between arms.

Inferential Analyses:
To determine if absenteeism differs between participants given COVID-19 educational material and the participants in the control group, we will conduct a complete-case analysis to compare mean absences between the intervention and control groups using a two-sample t-test.

We hypothesis implementing three COVID-19 education comic books for children and two COVID-19 education videos for parents will lower child's absenteeism and COVID infection rates and increased vaccine uptake.

School days missed will be modelled by a mixed zero inflated negative binomial model, with school specific random effects to account for clustering. COVID positive testing results will be modelled by a mixed logistic regression model, with school specific random effects to account for clustering.

Vaccine uptake will also be analyzed by mixed logistic regression model.

iii. Secondary outcomes

The secondary outcomes for the trial include measures from the PROMIS – Pediatric and Parent Proxy Instruments. The PROMIS instruments used are Children's PA, sleep, stress, depression, and anxiety; Parents' PA, sleep, stress, depression and anxiety. We will compare the scores for each of the items between the control and intervention arms at months 5 and 7. We will use a linear mixed effects regression models to determine whether between-arm differences were statistically significant.

iv. Additional Analysis

A sub-analysis will be done on the intervention group to determine if levels of engagement with the educational materials have impact on the primary and secondary outcomes. Level of engagement will be determined by participants self-report of reading/viewing.

Additionally, we will look at the interaction between the intervention group and other population risk factors to see if there are other effect modifications.