Statistical Analysis Plan (SAP)
Adolescent and child Suicide Prevention In Routine clinical Encounters (ASPIRE) study

Official Title: A comparative effectiveness trial of strategies to implement firearm safety promotion as a universal suicide prevention strategy in pediatric primary care

Funding Agency: National Institute of Mental Health (NIMH)

NIH Grant Number: R01 MH123491

University of Pennsylvania Protocol Number: 844327

Principal Investigator: Rinad S. Beidas, PhD

ClinicalTrials.gov Number: NCT04844021

Original SAP: 09 August 2021

Final SAP: 03 July 2023

Table of Contents

Final Statistical Analysis Plan

Date of approval by study team: 03 July 2023 (Prior to unblinding and data analyses; primary outcome data analyses commenced on 25 July 2023)

ASPIRE Statistical Analysis Plan for Primary Comparison of Nudge vs Nudge+

Primary outcome: Reach, defined as the composite (multiplicative) of two binary outcomes:
1) discussion of safe firearm storage (Yes/No)
2) offered a cable lock (Yes/No)

Therefore, the primary outcome will be coded as a 1 (successful program delivery) if both components were documented as "Yes" in the EHR; otherwise, the primary outcome will take value 0. The primary outcome is measured at the patient-visit level. If the same patient has multiple visits data from each visit will be included in the analysis. If multiple children in the same family, all visits for all children will be included.

As a secondary analysis, we will analyze 1) and 2) individually. We hypothesize that the analysis of component 2 (cable lock offers) will provide results similar to the results for the composite Reach outcome since fewer clinicians will endorse the cable lock than discussion and it is rare for clinicians to offer the lock without also doing the discussion.

Missingness in the primary outcome: Due to the nature of the EHR systems at the two sites, there is a non-negligible amount of missingness in the two binary outcomes that make up the composite Reach variable. More specifically, there is a large amount of missingness in the "offered a cable lock" outcome at one site, since clinicians were able to exit the EHR smartlist after documenting the first outcome about safe storage discussion. Unfortunately, missingness in one of the two outcomes induces missingness in the composite Reach outcome. In the primary analysis, we will treat any missing value as a "No." As a result, there will be no missing values in the Reach outcome for the primary analysis.

Statistical analysis plan:

Here, we describe the statistical analysis plan for the primary analysis of Reach.

Check balance across arms:
We will first calculate standardized mean differences (SMD) of both individual-level (sex, race, etc.) and clinic-level (size, presence of stakeholder, etc.) variables. Variables with a SMD greater than 0.2 will be included as covariates in the outcome models. The SMD for a given variable will be calculated as:

$$SMD = \frac{\bar{Y}_{2\cdot} - \bar{Y}_{1\cdot}}{\hat{\sigma}_{Pooled}}$$

where $\bar{Y}_{2\cdot}$ is the sample average of Nudge+ cluster means; $\bar{Y}_{2\cdot}$ is the sample average of Nudge cluster means; and $\hat{\sigma}_{Pooled}$ is the standard deviation of all N=30 cluster means.

Primary outcome model:
Missing values in the Reach outcome will be imputed as having not delivered the S.A.F.E. Firearm program. Given the definition of Reach will be completely observed (no missingness), we plan to model the patient-level binary Reach outcome using a logistic generalized estimating equation (GEE) with an exchangeable correlation at the clinic level. This model provides an estimate of the marginal effect of Nudge+ versus Nudge, i.e., the expected difference in the population if all clinics had received Nudge+ compared to giving all clinics Nudge only. In addition to the clinic-level indicator of randomized treatment assignment (corresponding to the parameter of interest), we will include the following set of covariates: sex (patient-level), health system (clinic-level), clinic size measured by number of well-child visits in the year prior to the trial launch (clinic-level), and clinician partner status (clinic-

level). We chose to include patient-level sex based on pilot findings that the rate of program delivery differs significantly for male versus female patients and by patient race. The three clinic-level covariates will be included because they were used in the covariate constrained randomization and are therefore anticipated to be highly related to Reach.

Because of the moderately small number of clinics in the trial, we will apply the Kauermann and Carroll small cluster correction to avoid underestimation of variance (e.g., using EMPIRICAL=ROOT statement with PROC GLIMMIX in SAS). Furthermore, if effort and availability of reliable software allow, we will use a custom nested exchangeable correlation instead of an exchangeable correlation clustering at the clinic level. The nested exchangeable structure will contain two parameters: one for a common correlation within clinician and the other for the correlation between different clinicians practicing at the same clinic. If available, the nested exchangeable correlation structure be more efficient if it more closely resembles the true correlation structure in the data.

Hypothesis testing:

We have proposed to use a three-sided hypothesis test using a Type I error of 0.05 to simultaneously test whether 1) Nudge+ and Nudge are equivalently effective, and 2) Nudge+ is superior to Nudge. This hypothesis testing framework required our statistical team to elicit an equivalence margin from the study investigators. To prioritize interpretability and actionability of our main study findings, we elicited the margin on the marginal probability scale. We determined that a margin of 10 percentage points would be reasonable to declare the two interventions equivalent. In other words, if we determine with statistical confidence that the rate of Reach in Nudge+ is more than 10 percentage points higher than the rate in Nudge, then Nudge+ will be deemed superior. A three-sided testing framework allows us to test equivalence, inferiority, and superiority without adjusting the Type I error for multiple tests. To implement this framework and perform these tests on the marginal probability scale, we will construct a series of Wald tests for the risk difference along a grid of null parameter values. The risk difference and its standard error (SE) will be obtained using PROC GLIMMIX's MARGINS statement. The SE will reflect the small cluster correction to the degrees of freedom (DF) by the EMPIRICAL statement (we will check that that the degrees of freedom match). For a given null value, μ_s, we will compute the t-statistic:

$$t_s = \frac{EST - \mu_s}{SE}$$

where EST denotes the estimated risk difference. We will compare this t-statistic to the critical value of the t-distribution with appropriate DF to obtain a p-value and reject the null hypothesis that the risk difference is equal to μ_s. We will repeat this process for a fine grid of μ_s values, $s \in \{1, \dots, S\}$, and form a confidence interval for the true risk difference. If the confidence interval is completely contained within the equivalence margin, Nudge and Nudge+ will be concluded to be equivalently effective. If the lower limit of the confidence interval is greater than the upper limit of the equivalence margin, Nudge+ will be deemed superior to Nudge.

SAP amendments

1. Details were added to further operationalize the analysis plan from what was originally specified in the study's protocol paper (Beidas et al., 2021).

2. The primary outcome of the study was changed from parent-reported receipt of *S.A.F.E. Firearm* (i.e., fidelity) to clinician-reported delivery of *S.A.F.E. Firearm* in the electronic health record (i.e., reach).
    - Both the study's National Institute of Mental Health Program Officer and Data Safety and Monitoring Board approved of this change on 02 March 2022, ahead of the trial's launch on 14 March 2022.
    - A detailed description of the rationale for this change is below as submitted to funding agency, National Institute of Mental Health.

Summary. Insights from behavioral economics, or how individuals' decisions and behaviors are shaped by finite cognitive resources (e.g., time, attention) and mental heuristics, have been underused in efforts to increase the use of evidence-based practices in implementation science. Using the example of firearm safety promotion in pediatric primary care, which addresses an evidence-to-practice gap in universal suicide prevention, we aim to determine: is a less costly and more scalable behavioral economic-informed implementation strategy (i.e., "Nudge") powerful enough to change clinician behavior or is a more intensive and expensive facilitation strategy needed to overcome implementation barriers? The Adolescent and child Suicide Prevention in Routine clinical Encounters (ASPIRE) hybrid type III effectiveness-implementation trial uses a longitudinal cluster randomized design. We will test the comparative effectiveness of two implementation strategies to support clinicians' use of an evidence-based firearm safety practice, *S.A.F.E. Firearm*, in 32 pediatric practices across two health systems. All pediatric practices in the two health systems will receive *S.A.F.E. Firearm* materials, including training and cable locks. Half of the practices ($k = 16$) will be randomized to receive Nudge; the other half ($k = 16$) will be randomized to receive Nudge plus 1 year of facilitation to target additional practice and clinician implementation barriers (Nudge+). The primary implementation outcome is parent-reported clinician fidelity to the *S.A.F.E Firearm* program. Secondary implementation outcomes include reach and cost. To understand how the implementation strategies work, the primary mechanism to be tested is practice adaptive reserve, a self-report practice-level measure that includes relationship infrastructure, facilitative leadership, sense-making, teamwork, work environment, and culture of learning. The ASPIRE trial will integrate implementation science and behavioral economic approaches to advance our understanding of methods for implementing evidence-based firearm safety promotion practices in pediatric primary care. The study answers a question at the heart of many practice change efforts: which strategies are sufficient to support change, and why? Results of the trial will offer valuable insights into how best to implement evidence-based practices that address sensitive health matters in pediatric primary care.

Description of problem. Parental response to our brief survey, the method through which we collect our primary implementation outcome of fidelity, has been lower than anticipated in the pilot and pre-implementation period. Our health system partners report that overall survey response in the health systems is much lower across all projects since COVID-19. The major methodological threats of a low response rate include: (a) concerns about representativeness; (b) lower power; and (c) the need to impute missing data for most of the sample. Given these concerns, our DSMB methodologist, Danny Almirall, PhD, recommended that we switch our primary outcome to reach as measured by physician electronic medical record documentation.

Steps taken. We began piloting the parent survey in May 2021 to maximize parental response rate. Our initial plan, as proposed in the grant submission, was to engage in a combination of email, text message, and mailed letters (i.e., methods found to be effective in previous work in these health systems). From May 2021 to February 2022, we conducted 13 pilots to maximize response rate, systematically changing how we approached collecting this data to ascertain the most effective approach (e.g., changing the order of outreach methods; adding messaging through patient portals, adding phone calls). Using a range of approaches, we obtained a 23% response rate, with a strong signal that phone calls are the most effective manner to obtain parental responses.

Given our desire to keep fidelity as our primary outcome, we tried two additional approaches to increase response rate. First, we piloted sending a one-question survey via email and/or text to reduce the friction and time commitment associated with responding. This resulted in a 10% response rate and is likely due to common carriers blocking text

messages that include 'firearm,' even though we tried many strategies to circumvent these blocks. Furthermore, it resulted in loss of valuable data since only 1 question could be asked.

Second, we tried an intensive outreach approach that included 2+ phone calls to a smaller subset of randomly selected patients (n = 100) per health system every 2 weeks. This approach allowed us to reach a 37% response rate, which is the highest response rate we have been able to obtain.

The intensive outreach approach is the most effective one. Unfortunately, the sites are not resourced from a staff perspective to engage in this approach. Making phone-calls to 100 parents every 2 weeks per site requires hundreds of phone calls and places substantial strain on our health system partners with their current staffing model. We have explored budget reallocations but we will not be able to reallocate sufficient additional resources to the sites to support the additional staff needed for this resource intensive approach, given initial budgetary cuts and unanticipated increases to the cost of gun locks due to supply chain issues.

Proposed change. We propose to change the primary outcome in the ASPIRE trial from fidelity, measured via parental report, to reach, measured via physician documentation in the EHR. Our health systems have built integrated templates in the electronic medical record for the physicians as part of the trial that enable seamless documentation and efficient data extraction. We will continue to collect fidelity as a secondary outcome from a subset of participants.

Advantages of this approach. First, we will have a complete dataset with no missing data, thus strengthening the rigor of our conclusions. Second, while we acknowledge reach will likely be an overestimate of what is occurring, the bias will impact the two arms in the same way (see table below). Third, we will still collect fidelity from a subset of participants. Our partners are currently calculating the number of patients per two weeks that they can outreach to using the intensive outreach approach. We will work with our biostatistician, Dr. Kristin Linn, to potentially adjust for misclassification bias in our reach estimates by using the smaller, gold-standard validation sample of fidelity data.

Comparison of reach and fidelity (October 21-Feb 22)		
	EHR – Reach	**Survey – Fidelity**
Counseling	56.3%	47.9%
Locks	46.4%	31.8%
Both	45.3%	30.0%

Original Statistical Analysis Plan

Date of approval by study team: 09 August 2021 (Prior to trial launch on 14 March 2022)

The original statistical analysis plan is described in the study's protocol paper:

Beidas RS, Ahmedani BK, Linn KA, et al. Study protocol for a type III hybrid effectiveness-implementation trial of strategies to implement firearm safety promotion as a universal suicide prevention strategy in pediatric primary care. Implement Sci 2021;16.

Study blinding details

The study's lead biostatistician (KAL) generated the randomization scheme and assigned clinics to study arm. The research team at each health system (Site PI JMB and study staff LQ and LW for Colorado site; Site PI BKA and study staff CP for Michigan site) enrolled clinics.

Lead biostatistician KAL generated the randomization scheme and allocated clinics (by name) to study condition. The clinics were assigned a random ID number and study arms were also assigned a random ID number by author CJ. Study arm and clinic names were not used when discussing study activities with the blinded individuals throughout the trial. The study's Principal Investigator and this manuscript's first author (RSB), the biostatistician team (KAL, SCM), and the study's data analyst (KB) were blinded to clinic assignment at the time of data analysis. The ID numbers were listed in lieu of clinic and study arm names in the datasets used for analysis. The blinded individuals were not unblinded to clinic names and study arm until after analyses had been completed and manuscript writing commenced.