**SAMPLE**

**4.4 Statistical Design and Power**

**Outcomes**

Our primary outcome is the receipt of vaccine in adolescents between the intervention and control periods. As a secondary way of looking at receipt of vaccine, missed opportunities is one inverse calculation of receipt. This can be defined as patient visits for any reason, including but not restricted to well visits, in which a patient was due for vaccine, but in which no vaccine was administered. Measuring missed opportunities will be one way of assessing improved receipt of vaccine. Receipt of vaccines and missed opportunities will be assessed through medical record review of randomly chosen adolescent visits in each practice. Others factors that may affect receipt of vaccine, including reduced vaccine hesitancy and increased uptake of adolescent well visits will also be measured for each practice.

Our secondary outcome, assessed in Aim 2, is receipt of the subsequent dose of vaccine. As we review medical records for missed opportunities, we will also be identifying visits that were visits where vaccine was due and delivered to the adolescent (captured opportunity). For those captured opportunities where another dose of vaccine is required, we will follow these patients for 12 months and subsequent doses of vaccine captured.

Additional outcome measures include visit experience and visit satisfaction. We will also evaluate patterns of communication and associations with practice environment, demographic and socioeconomic factors, and English proficiency.

**Number of Subjects, Effect Size, and Power**

All sample size estimates assume a two-side alpha of .05. The project’s planned sample size was powered to address Specific Aim 1, whose outcome is the decrease in the proportion of office visits involving adolescents due for vaccine that result in missed opportunities for vaccination. In each of 36 practices in the stepped wedge design, 30 eligible records in each practice are reviewed on average 3.5 times before and 3.5 times after initiation of the communication strategy (see Research Strategy, Figure 2).We calculated the sample size estimate and power by first determining the required sample size for a two-group comparison of proportions between intervention and control and then inflated this sample size to account for correlation within practices over time according to the stepped wedge design by applying a design effect (see Research Strategy for details).

This sample size is based on an intra-cluster correlation (ICC) of 0.14 estimated from preliminary studies and will afford power of 90% to detect a reduction in the percentage of visits that result in missed opportunities from 50% before intervention to 44% (6 percentage point reduction) after initiation of the communication strategy.

Figure 2 in the Research Strategy illustrates the timing of measurements in the proposed design, which initiate the intervention in 36 clinics over four steps. The number of pre- and post-intervention measures depends on the step. Because many clinics will be measured more frequently (Figure 2), estimates of power are also conservative.

For Aim 2, assuming that 50% of the adolescents surveyed in Aim 1 receive the vaccine and that 75% of those will require a 2nd dose, we estimate that we will follow 594 adolescents in the control period and 648 in the intervention period for at least 12 months for receipt of the 2nd vaccine dose. Using the same methods as for Aim 1 but assuming an ICC of 0.1. If 50% of the adolescents in the control period receive the 2nd vaccine, this sample size affords 80 percent power to detect a between-group difference in proportions of 0.07, or a 7 percentage point difference.

**Statistical Method**

Aim 1 willcompare effectiveness of our educational intervention on recommendation strategiesvs. the control of usual care in decreasing missed opportunities for vaccination (Primary Outcome). *Hypothesis: Adolescents seen in practices trained in the recommendation strategy will demonstrate a larger decrease in missed opportunities than those practices when in the usual care control.* Missed opportunities will be measured initially and then every 6 months in each practice. A missed opportunity is a patient visit for any reason in which vaccine is due and is not administered. Our design is a stepped wedge cluster randomized trial (SW-CRT). Our basic analytic model will be a generalized linear mixed effects model. This will include using generalized linear models to account for clustering by practice and facilitator (random effects), a time variable to account for secular trends and stratification by cluster if needed to satisfy modeling assumptions. To enhance the interpretability of results, we will explore logistic, log binomial and log linear regression models that generate ratios of the odds, ratios of the probabilities, and differences in the probabilities, respectively, that an opportunity is missed.

The data we analyze in Aim 1 will consist of seven independent counts and proportions of missed opportunities in each of the 36 clinics. A statistical model that assesses the time\*arm interaction can determine if the volume of missed opportunities over time is smaller among clinics in the intervention period than in the control period accounting for temporal variables. We will explore the effect of practice type, sex and race/ethnicity of the adolescent by including interactions between these factors and intervention/control.

Aim 2 will assess the comparative impact of the intervention on receipt of a subsequent dose of vaccine (Secondary Outcome). *Hypothesis: Adolescents seen in practices trained in recommendation strategies will more often accept subsequent vaccine doses than adolescents who are seen when the practices are in usual care (control).*

The study design permits identification of two groups, numbering 594 and 648 adolescents in the control and intervention period respectively, who will receive a 1st and need a second dose and will be followed for at least 12 months and compared on the proportion who receive subsequent vaccination.

The basic analytic plan will mirror Aim 1. We will use multivariable logistic and binomial regression models to compare subsequent vaccination between the control and intervention periods, and the proposed design can detect even smaller differences in proportions. The model will also accommodate within-practice clustering by treating practice as a random effect.

Aim 3 will includeaddress the additional outcomes through process evaluation.To characterize parent-provider and office-wide communication practices involving vaccine beyond recommendations; we will do office structured environmental scans; assess the process of visits with interviews with providers, nurses and office staff; and conduct qualitative discourse analysis. *Intention: To demonstrate how communication practices and environmental factors characterize the delivery of vaccines to adolescents.*  Rather than a hypothesis, a statement of intention conforms to conventions of qualitative data.

The power of discourse analysis is not statistical, but descriptive and observational, in its ability to identify patterns, positions speakers take, assumptions, preferences, perspectives, and ways they make and interpret meaning. In the proposed analysis, two coders will rate each transcript using the Four Habits Coding Scheme (FHCS), a validated instrument that rates provider communication behaviors and provides a score that will be compared with discourse analysis of the transcripts that codes the speakers ways of using language as described above for patterns that may explain the score. This approach may generate a hypothesis for future testing**,** highlighting effects related to the communication style on vaccine acceptance. Informed consents will ask willingness to have de-identified transcripts used in future research, thus creating a valuable communication database as an asset to additional pediatric communication analyses.

The sample will be 4 family-provider dyads chosen at random from the 12 audio-recorded visits in each of the 36 practices for a total of 144 dyads, to assure representative patterns and provide audit and feedback during the intervention and 6 months afterward. Discourse analysis keyed to positioning theory will be used for initial interpretation of full interactions, triangulated by computer-assisted qualitative content analysis, in which the initial coding of language categories is automated by WMatrix®, a program which disambiguates and identifies words falling into semantic clusters or domains in a text. The data-driven categories are then interpreted by the researcher, first using discourse analysis to code stance, participation and speech practices and then triangulated by log-likelihood calculation of significance. Excerpts from recordings showing vaccine recommendations will be coded and interpreted separately, Field notes taken during the environmental structured scans will also be analyzed as an external view of vaccine practice, a procedure often used in medical anthropology. Though lacking the power of statistical analysis, this approach to process evaluation provides triangulation of the patterns that discourse analysis detects to explain findings in Aim 1 and Aim 2.

**Missing data**

Data from all sources will be collected by project research staff, either the practice facilitator during scheduled visits or a research assistant, in participating practices. All research staff will be trained by PIs and overseen by the network directors for consistency. To preserve data integrity, individual practices will not collect their own data. We do not anticipate barriers to obtaining the number of medical records, family and provider surveys, or encounter recordings that we have specified. For surveys missing data patterns will be summarized and multiple imputation will use chained equations.

*Date: 03/18/19*

*Contact: Rebecka Bourn, PhD, OSCTR Science Writing Unit*

*OSCTR Website:* [*http://osctr.ouhsc.edu*](http://osctr.ouhsc.edu)

*Funding: NIGMS award U54GM104934*