

Increasing Teen Access to Care (AccessKCTeen2)

NCT05444855

05 September 2023

Statistical analysis plan

We will use R software (R Foundation for Statistical Computing) for analyses. We will conduct descriptive analysis with inferential and graphical exploratory analytic techniques, presenting mean, median, and interquartile range [continuous variables] and frequency and percentage [categorical variables]. Next, bivariate analyses will be conducted to assess the relationship between variables. We will examine data for outliers and/or high leverage points and use established best practices in decided how to address them.

We will compare outcomes using independent t-tests for continuous variables and Chi-Square or Fisher's exact test for categorical variables depending on cell counts. If continuous variables are highly skewed nonparametric testing will be implemented (i.e., Wilcoxon rank-sum test). All tests will be two-sided, assessed with significance of $p<0.05$. We will evaluate these feasibility constructs:(acceptability, demand, implementation, practicality, integration, expansion, limited efficacy [see primary outcome]) using descriptive statistics. *Primary outcome:* We will calculate the proportion of adolescents with any health care utilization at 6 months.

Attrition: We will use established methods to minimize attrition (see "Recruitment and Retention") and plan to assess for our primary outcome every two months to keep participants engaged and minimize bias. Participants who are lost to follow-up after three attempts at 6 months will be excluded from the denominators for hypothesis testing. We will report on demographics of those lost to follow-up compared to those who completed the study.

Missing data: We plan to minimize missing data by engaging our CAB to pretest survey measures to achieve brevity and clarity. Our electronic data collection tool (REDCap) can be set to require participant responses before advancing. In order to identify potential problems promptly, we will run a REDCap report every second week to review the degree of missingness for each variable. Any variable missing $>5\%$ will be reviewed to determine probable causes and a plan will be developed to address the issue(s). During analysis, we will determine both the extent of missingness and whether data are missing at random (MAR). If the data are missing for $<20\%$ of responses, we will rely on list-wise deletion. If $\geq 20\%$ is missing we will utilize multiple imputation techniques. Specifically, we will determine if data are MAR by partitioning the data into two groups (those with and without missing data) and examining the distributions across variables of interest and employ appropriate imputation techniques.