Statistical Analysis Plan

Title: Sustaining Patient-centered Alcohol-related Care (SPARC)

NCT: NCT02675777

Dated: 7/13/2018

Last updated: 8/25/2022

Statistical Analysis Plan for SPARC Trial

Background

Alcohol use is a major cause of disability and death worldwide. To improve prevention and treatment addressing unhealthy alcohol use, experts recommend that alcohol-related care be integrated into primary care (PC). However, few healthcare systems do so. To address this gap, implementation researchers and clinical leaders at Kaiser Permanente Washington (KPWA) partnered to design a high-quality Program of Sustained Patient-centered Alcohol-related Care (SPARC).

The study aim is to evaluate the effectiveness of the SPARC implementation at improving alcohol-related care within 22 KPWA clinics as part of a stepped-wedge randomized implementation trial. Specifically, our primary objectives are to evaluate whether the SPARC implementation intervention increased (1) brief interventions for unhealthy alcohol use, and (2) initiation and engagement in treatment of alcohol use disorders (AUDs). Details on the specific components of the SPARC implementation are in the protocol paper (Glass et al. 2018).

Trial registration: NCT02675777

Stepped-wedge trial design

To evaluate the effectiveness of the SPARC implementation, we are conducting a pragmatic stepped wedge trial in 22 KPWA clinics. The trial design (**Figure 1**) addressed several practical concerns of the health system while maintaining a rigorous evaluation of SPARC. The 22 clinics were randomized into 7 mutually exclusive groups of clinics (referred to as "waves"). Here we briefly summarize the features of the design; more detailed explanations for these requirements are described in the protocol paper (Glass et al. 2018):

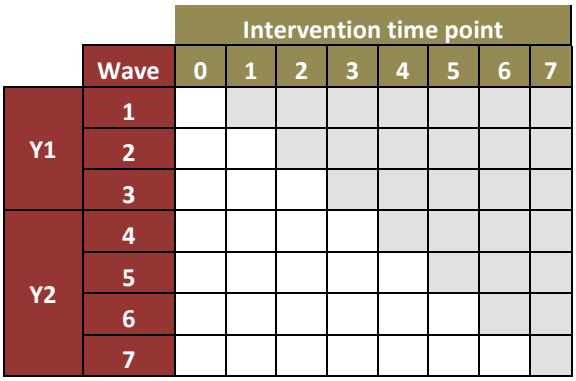


Figure 1: Schematic illustrating the stepped wedge design, in which 19 sites were divided into 7 waves (2-3 sites per wave) and stratified across two groups (Y1 and Y2). Gray squares denote time periods in which the SPARC implementation was to be implemented based on randomization, which includes the active implementation period and the sustainment period (defined below).

- A few pairs of clinics were clustered together into larger single clinical sites for the purpose of receiving the SPARC implementation. This resulted in 19 distinct clinical sites, hereafter referred to as "sites."
- The health system provided input on implementation timing by conducting a stratified randomization on the year of the implementation; the health system identified 9 sites that would be randomized in Year 1 (3 sites in each of 3 waves), and the remaining 10 sites that would be randomized in Years 2-3 (in two waves of 3 sites each and two waves of 2 sites each). We will refer to sites randomized in Year 1 as Y1 sites and sites randomized in Years 2-3 as Y2 sites.
- It was desired that one of the Y2 sites be in a wave with just a single other site
- It was desired that the final wave consist of only 2 sites.

For each wave, there was a 2-month preparatory period prior to the "launch date" of the 4-month period of active implementation. The "launch date" was specified at the time of randomization to study

wave. It was the date when each site was intended to start Behavioral Health Integration screening for all patients who came to the site.

Randomization

As discussed above, randomization was to be stratified on the study year of implementation (Y1 vs. Y2). The 9 Y1 sites were randomly assigned to begin active implementation in one of 3 waves (3 sites per wave). Within Y2, the 4 implementation waves were to consist of 2 waves with 3 sites each, and 2 waves with 2 sites each. Given the requirements above that the last wave have just two sites and that one Y2 site needed to be in a wave with just one other site (we will refer to this site as site A), we developed a randomization scheme to ensure that each site had the same probability of being assigned to each wave (0.2 probability to be assigned to Y2 wave 4 and 0.8/3 each for waves 1-3 of Y2) as follows. First, we randomly selected one of the first 3 waves to be the other 2-site wave. Then, we randomly assigned site A to wave 4 with probability 0.2 or to the other 2-site wave with probability 0.8. The remaining 9 Y2 sites were then to be randomly assigned to the other available slots. Under this randomization scheme, letting B denote an arbitrary site different from site A, we have:

- P(site A assigned to wave 4) = 0.2
- P(site B assigned to wave 4) = P(site B assigned to wave 4 | A assigned to wave 4) * P(A assigned to wave 4) + P(site B assigned to wave 4 | A not assigned to wave 4) = 1/9*0.2 + 2/9*0.8 = 0.2
- P(site A assigned to wave *j*) = P(wave *j* assigned to be the other 2-site wave) * P(site A assigned to wave *j* | wave *j* assigned to be the other 2-site wave) = 1/3*0.8, for *j* = 1,2,3
- P(site B assigned to wave *j*) = 1/3*[1 – P(site B assigned to wave 4)] = 0.8/3, for *j* = 1,2,3

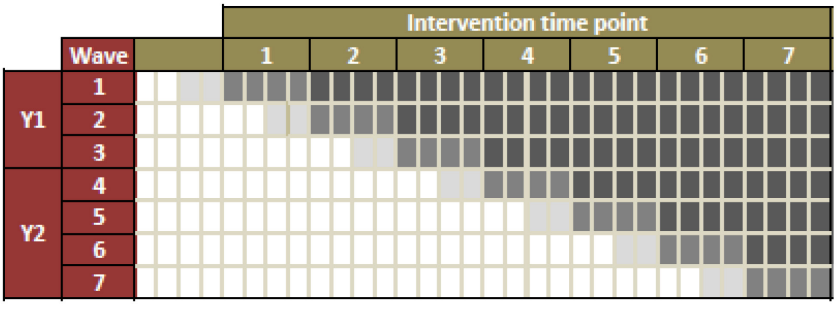
Allocation concealment

The random allocation sequences were generated by the study biostatistician (Dr. Bobb) after all of the KPWA sites were identified, including the pairing of clinics within 3 of the sites. Y1 sites were randomized to waves on January 22, 2016 and Y2 sites were randomized on October 7, 2016. The assigned waves are shown in **Table 1** below.

Study sample

Our base study population consists of patients who had at least 1 visit to one of the 22 KPWA clinics during the period from January 1, 2015 to July 31, 2018, which consists of the period from approximately 1 year prior to the randomization date of the Y1 sites (January 22, 2016) through the end of active implementation in the last wave of sites.

SPARC Implementation stages



Key	Period	Comparison (Primary)
	Pre Preparatory	Pre
	Active Sustainment	Post

Figure 2: Time windows for SPARC implementation

Implementation time windows of interest are shown in **Figure 2**. The "preparatory period" is defined as the period 2 months prior to the SPARC launch date up until the launch, and the "active

implementation" period is defined as the period from 1-4 months after the SPARC launch date at a given site. We will refer to the period up to and including the preparatory period as the "pre-implementation" period, and the period starting with the launch date as the "post-implementation" period. Following intention-to-treat principles, unless otherwise specified, these pre- and post-implementation periods will be defined based on the launch dates during which active implementation was planned to occur (shown in **Table 1**), rather than the actual date when the site begins implementing SPARC (if implementation was delayed).

Table 1. Implementation wave dates and randomized assignments of sites to waves

Year	Wave	Site (masked)	Launch Date
Y1	1	Site A	April 4th, 2016
Y1	1	Site B	April 11th, 2016
Y1	1	Site C	April 18th, 2016
Y1	2	Site D	September 7th, 2016
Y1	2	Site E	September 12th, 2016
Y1	2	Site F	September 19th, 2016
Y1	3	Site G	January 4rd, 2017
Y1	3	Site H	January 9th, 2017
Y1	3	Site I	January 16th, 2017
Y2	4	Site J	April 24th, 2017
Y2	4	Site K	April 24th, 2017
Y2	4	Site L	April 24th, 2017
Y2	5	Site M	August 14th, 2017
Y2	5	Site N	August 14th, 2017
Y2	5	Site O	August 14th, 2017
Y2	6	Site P	December 4th, 2017
Y2	6	Site Q	December 4th, 2017
Y2	7	Site R	April 9th, 2018
Y2	7	Site S	April 9th, 2018

Outcome measures and time frames

The main primary and secondary outcome measures, as described in the Protocol Paper, are defined in **Table 2** below. Indicator variables for each outcome are created at the level of the primary care visit (since a patient may have multiple primary care visits during the study period). Measures are then aggregated into time-intervals (e.g., weekly or monthly). For the primary analysis, measures are aggregated into 28-day intervals (hereafter referred to as "months"), defined relative to the official launch date of SPARC within the clinic (denoted by T0). For example, the co-primary alcohol brief intervention measure aggregated at the monthly level is defined as the proportion of patients seen in the month who had a brief intervention (defined in **Table 2**) at any point during that month.

Table 2. SPARC Trial Primary, Secondary and Other Outcomes from EHR and Claims Data

Category	Measure	Description
Primary Outcomes		

Prevention	Alcohol Brief Intervention	Indicator for whether a patient had a brief intervention documented in the EHR* on the day of, or in the 14 days following a PC visit, and had a positive alcohol screen on the day of the visit or in the prior 365 days*
Treatment	Treatment for Newly Diagnosed AUD (NCQA 2017 alcohol and drug treatment measure)	Indicator for whether a patient had a new AUD diagnosis* and initiated and engaged in AUD treatment*
Intermediate Outcomes		
Prevention	Alcohol screening documented	Indicator for whether a patient had AUDIT-C screening documented in the EHR on the day of the visit or in the prior 365 days
Prevention	Positive alcohol screen	Indicator for whether a patient screened positive on the AUDIT-C (3–12 women and 4–12 men), on the most recent screen documented on the day of the visit or in the prior 365 days
Prevention	High positive alcohol screen	Indicator for whether a patient had a high-positive AUDIT-C score (7–12 points), on the most recent screen documented on the day of the visit or in the prior 365 days
Assessment	Assessed for DSM-5 AUD symptoms	Indicator for whether a patient with a high-positive screen in the past year completed an AUD Symptom Checklist on the day of the visit or in the prior 365 days
Identification	Past-year AUD diagnosis	Indicator for whether a patient had an AUD diagnosis defined as an ICD code for an AUD diagnosis per NCQA anywhere in or outside KPWA (e.g. includes claims) on the day of the PC visit or in the prior 365 days
Identification	New AUD diagnosis	Indicator that a "past-year AUD diagnosis" (defined immediately above) was new on the day of the PC visit, based on no AUD diagnosis in the prior 365 days
Treatment	Initiation of AUD treatment (NCQA)	Indicator for whether a patient received a "new AUD diagnosis" (defined above) and initiated AUD treatment in the following 14 days, per HEDIS ICD codes
Treatment	Engagement in AUD treatment (NCQA)	Indicator for whether a patient who initiated AUD treatment (defined above) had another 2 treatment visits in the following 30 days after initiation ("engagement") per HEDIS ICD codes

* Definitions based on intermediate outcomes;
EHR, electronic health record; HEDIS, Healthcare Effectiveness Data and Information Set; NCQA, U.S. National Committee for Quality Assurance

For each outcome measure above, we will perform the following comparisons of the monthly outcome rates (see **Figure 2** for definitions of implementation periods):

Primary: before vs. after the start of the active implementation period

Secondary:

(1) active implementation vs. pre period
(2) sustainment period vs. pre period
(3) sustainment period vs. active implementation

ADDENDUM NOTE: We plan to report analyses of these secondary time periods in a second manuscript.

Power calculations

Using the method of Hussey and Hughes (2007), power was calculated based on 19 sites and 7 study waves (with the number of sites per wave as above), and with 4 months between waves. Calculations assumed that the average number of patients seen in a site during a month was 1,205 patients, based on 2014 data for these sites. We assumed the following pre-implementation rates for the main study outcomes based on KPWA screening data from 2012 and treatment data from 2014:

- Alcohol brief intervention outcome (defined as the proportion of primary care patients seen in the clinic who screened positive for unhealthy alcohol use and were given a brief intervention): 34.2 per 10,000 patients seen (0.342% = 19% screened x 36% screened positive x 5% brief intervention). Note that the estimated percentage of patients who receive brief intervention is unknown, and the value of 5% used here is thought to be an upper bound.
- NCQA treatment outcome (defined as the proportion of primary care patients seen in the clinic who initiated and engaged in treatment): 3.9 per 10,000 (0.039% = 1.26% newly diagnosed x 37.5% initiating treatment x 8.2% engaged).

To account for within-site correlation of patient outcomes, we further assumed a value for the intraclass correlation coefficient (ICC) of 0.001, equal to the observed ICC for AUD diagnosis rates across these sites for a 6-month period based on 2014 data. Note that these calculations assume that the ICC is constant over time, and that the correlation of patients from the same clinic at different time points is the same as for patients from the same clinic at the same time point (Hemming, Taljaard, and Forbes 2017). Power was calculated based on a two-sided test and a type 1 error rate of 0.05.

Based on these calculations, we expect to have 80% (90%) power to detect an increase in brief intervention rates of 7.1 (8.2) per 10,000 patients seen and an increase in the treatment outcome of 2.6 (3.1) per 10,000 patients seen. The following table shows power at other effect sizes:

Outcome	Assumed effect size: change in number (per 10,000 patients seen) before vs. after implementation	Power
Brief Intervention	6	0.67
	8	0.89
	10	0.97
Treatment	2	0.60
	3	0.89
	4	0.98

Descriptive analyses

Descriptive statistics

We will generate summary statistics on the number of visits per person to a site during a monthly period. At the patient level, we will describe demographic characteristics and clinical variables across patients with a visit to a clinic in the pre-implementation period and patients with a visit in the post implementation period. For time-varying disease status variables (e.g., the presence of a documented diagnosis), characteristics will be summarized based on whether the patient had the disease documented at any visit during the period (pre- or post-implementation), as well as based off of the patient's first visit during the period. Additionally, as suggested by the CONSORT extension for stepped wedge trials (Hemming et al. 2018), clinic and patient-level variables will be described by allocated sequence (i.e., study wave).

Crossover

We will examine the proportion of visits in which the patient visited a site during the pre-implementation period after having had a prior visit, to a different site during that site's post-implementation period (main measure). Additionally, as a general measure of crossover across sites, for each site we will compute the proportion of patients who had a visit at that site who also had a visit at each of the other sites. As a measure of crossover across the implementation periods, we will compute the proportion of patients who had a visit both before and after the launch date, overall and by site.

Graphical analysis of study outcomes

To visualize how the study outcomes change over time within a site, as well as pre- versus post-implementation, we will plot the site-specific, crude (unadjusted) rates of the study outcomes as a function of study month.

Primary analysis

Following the general framework for analyzing data from a stepped-wedge trial (Hussey and Hughes 2007; Hughes et al. 2015), for each outcome described above we will apply the following logistic mixed-effect model (GLMM),

$$logit\ P(Y_{ijm} = 1) = \alpha + \beta Int_{jm} + \gamma S_j + f(cm) + b_j + u_i,$$

where Y_{ijm} is the outcome for person *i* who visited site *j* in time interval *m*. As described above, the intervals used in analyses are 28-day periods ("months" hereafter) before and after the launch date for each clinic. The term Int_{jm} is an indicator variable for whether site *j* was in the post- versus pre-implementation period as of that month. The term S_j is an indicator for whether site *j* was a Y2 site versus a Y1 site (stratification variable), which accounts for possible differences in the outcome across these two years of sites (Y1 and Y2), and $f(cm)$ is a pre-specified function of calendar month to account for the potential for a secular trend in outcome rates over time. We plan to model $f(cm)$ using indicator variables for each 4-month calendar period. Alternate specifications will be considered in a sensitivity analysis (details below). Additionally, $b_j \sim N(0, \tau_b^2)$ is a site-level random effect to account for correlation of individuals from the same site and $u_i \sim N(0, \tau_u^2)$ is a person-level random effect to account for correlation of outcomes from the same individual, since a person can have a visit at multiple sites in the same or different month, or at the same site over multiple months.

The primary comparison (1) described above will be evaluated by first conducting a two-sided Wald test (at the 0.05 level) of the coefficient β. Ninety-five percent Wald confidence intervals (95% CI) for β will also be calculated.

For descriptive purposes, from the model, we will estimate the marginal predicted probability of the outcome in the pre- and post-implementation period by averaging over covariate distribution and the random effects.

To test the contrasts for secondary time periods of interest, the term Int_{jm} will be replaced with a categorical variable for whether the month was in the pre-, active implementation, or sustainment period, and the relevant coefficient will be tested. For example, for testing whether there was a difference in the outcome rates in the active versus pre-implementation period, we will test the null hypothesis that the coefficient for the indicator for the active implementation period (versus the pre-period) is equal to zero. As above, from the model we will estimate the marginal predicted probability (and 95% CIs) of the outcome for each of the implementation periods.

Secondary analyses

Variations in the definition of the brief intervention outcome were developed to reflect a new NCQA alcohol screening and follow up measure.

We will repeat our primary analysis for the following alternate specifications of the main brief intervention outcome:

- Extend time window to 60 days to match NCQA
- Apply the NCQA definition of brief intervention, which includes codes irrespective of associated text and excludes brief interventions identified via natural language processing and extends the time window to 60 days.
- Only include brief interventions that were given in primary care clinics (the primary definition allows BIs documented in other visit types).

Variations in the definition of the AUD treatment outcome to reflect an updated NCQA measure and longer time windows which might be important in primary care.

We will repeat our primary analysis for the following alternate specifications of the main AUD treatment outcome:

- Expand the types of visits included to allow for telephone encounters per NCQA 2018 HEDIS measure
- Allow for a longer time window for treatment initiation and engagement as described in the Protocol (Glass et al. 2018).
- Use a more stringent measure that only allows visits to count as treatment if they are to a behavioral health provider, have a specialty addiction treatment code, or had a dispensed prescription for AUD medication active at the time of the initiation or engagement visit(s) in the same time windows as previously described from the new diagnosis (see Protocol, Glass et al. 2018)).

Missing data

Because the primary outcomes are defined based on the presence (or absence) of a recorded outcome event in the EHR, we do not have missing outcome information. We also do not have any missingness in covariates being adjusted for in the primary or secondary analyses (measures of implementation timing, stratification variable). Some of the descriptive variables may have missingness, such as race/ethnicity,

which may be recorded as "unknown." Missingness of any secondary outcomes or descriptive variables will be described.

One possible reason for a lack of documentation of an outcome could be due disenrollment from the health plan. To investigate this potential, we plan to examine the distribution of the number of days that a patient was enrolled within KPWA during the following time windows from the day of visit: the prior 365 days, the following 14 days, and the following 44 days.

References

Bobb JF, Lee AK, Lapham GT, Oliver M, Ludman E, Achtmeyer C, Parrish R, Caldeiro RM, Lozano P, Richards JE, Bradley KB (2017). Evaluation of a pilot implementation to integrate alcohol-related care within primary care. *International Journal of Environmental Research and Public Health.* 14(9). pii: E1030.

Glass JE, Bobb JF, Lee AK, Richards JE, Lapham GT, Ludman E, Achtmeyer C, Caldeiro RM, Parrish R, Williams EC, Lozano P, Bradley KA (2018). Study protocol: a cluster randomized trial implementing sustained patient-centered alcohol-related care (SPARC Trial). *Implementation Science*. 13:108.

Hemming K, Taljaard M, Forbes A (2017). Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. *Trials*. 18(1):101.

Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, Dixon-Woods M, Aldcroft A, Doussau A, Grayling M, Kristunas C, Goldstein CE, Campbell MK, Girling A, Eldridge S, Campbell MJ, Lilford R, Weijer C, Forbes A, Grimshaw, JM (2018). Reporting of The CONSORT extension for SteppedWedge Cluster Randomised Trials: Extension of the CONSORT 2010 statement with explanation and elaboration. BMJ. https://research.birmingham.ac.uk/portal/files/50005963/CONSORT_SW_accepted_version.pdf

Halekoh U, Højsgaard S (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models–the R package pbkrtest. *Journal of Statistical Software.* 59(9):1-30.

Hughes JP, Granston TS, Heagerty PJ (2015). Current Issues in the Design and Analysis of Stepped Wedge Trials. *Contemporary Clinical Trials*. 45:55-60.

Hussey A, Hughes JP (2007). "Design and Analysis of Stepped Wedge Cluster Randomized Trials." *Contemporary Clinical Trials* 28, no. 2: 182–91.

Kahan BC, Forbes G, Ali Y, Jairath V, Bremner S, Harhay MO, Hooper R, Wright N, Eldridge SM, Leyrat C. Increased risk of type I errors in cluster randomised trials with small or medium numbers of clusters: a review, reanalysis, and simulation study. Trials. 2016 Dec;17(1):438.

Li P, Redden DT. Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. BMC medical research methodology. 2015 Dec;15(1):38.