Statistical Analysis Plan

Mental Health Among Patients, Providers, and Staff (MHAPPS): Investigating the Mental Health Impact of COVID-19 and Comparing the Effectiveness of Two Caring Contact Interventions

Principal Investigator: Anna Radin, DrPH, MPH, Applied Research Scientist, St. Luke's Health System

Statistical Analysis Plan prepared by: Siobhan Brown, PhD (University of Washington)
Version 1.0, November 24, 2020

1 INTRODUCTION

MHAPPS is designed to assess to mental health impact of COVID-19 in both patients and employees/care providers in a large hospital system in Idaho (Saint Luke's Health System; SLHS). In addition to the pandemic, some of the mitigation measures used to contain spread, such as quarantines, shut-downs, and social distancing, may increase feelings of loneliness, anxiety, and depression. In the first stage of the study, the investigators will perform a cross-sectional survey to measure the prevalence of these and other measures of mental distress. In the second phase, survey participants who indicate moderate to high levels of one or more measures of mental distress will be invited to participate in a randomized trial comparing the effectiveness of two caring contact interventions. Caring Contacts is an evidence-based suicide prevention intervention, typically consisting of brief messages sent via letters, emails, or text messages from a provider to a patient. This study will use a text messaging platform to deliver Caring Contacts, sending 11 pre-scripted messages scheduled over a 6-month timeframe to the first intervention arm, and 11 caring text messages plus 1 call, within 2 weeks of enrollment to the second intervention arm.

2 SPECIFIC AIMS

Aim 1: Measure the prevalence of mental distress including loneliness, anxiety, depression, substance use, suicide ideation and other suicide-related risk factors in providers, staff, and patients served by SLHS. Describe differences associated with age, sex, race, ethnicity, urban or rural residence, living situation, occupation, local COVID-19 prevalence, perceived risk related to COVID-19, and masking and social distancing practices.

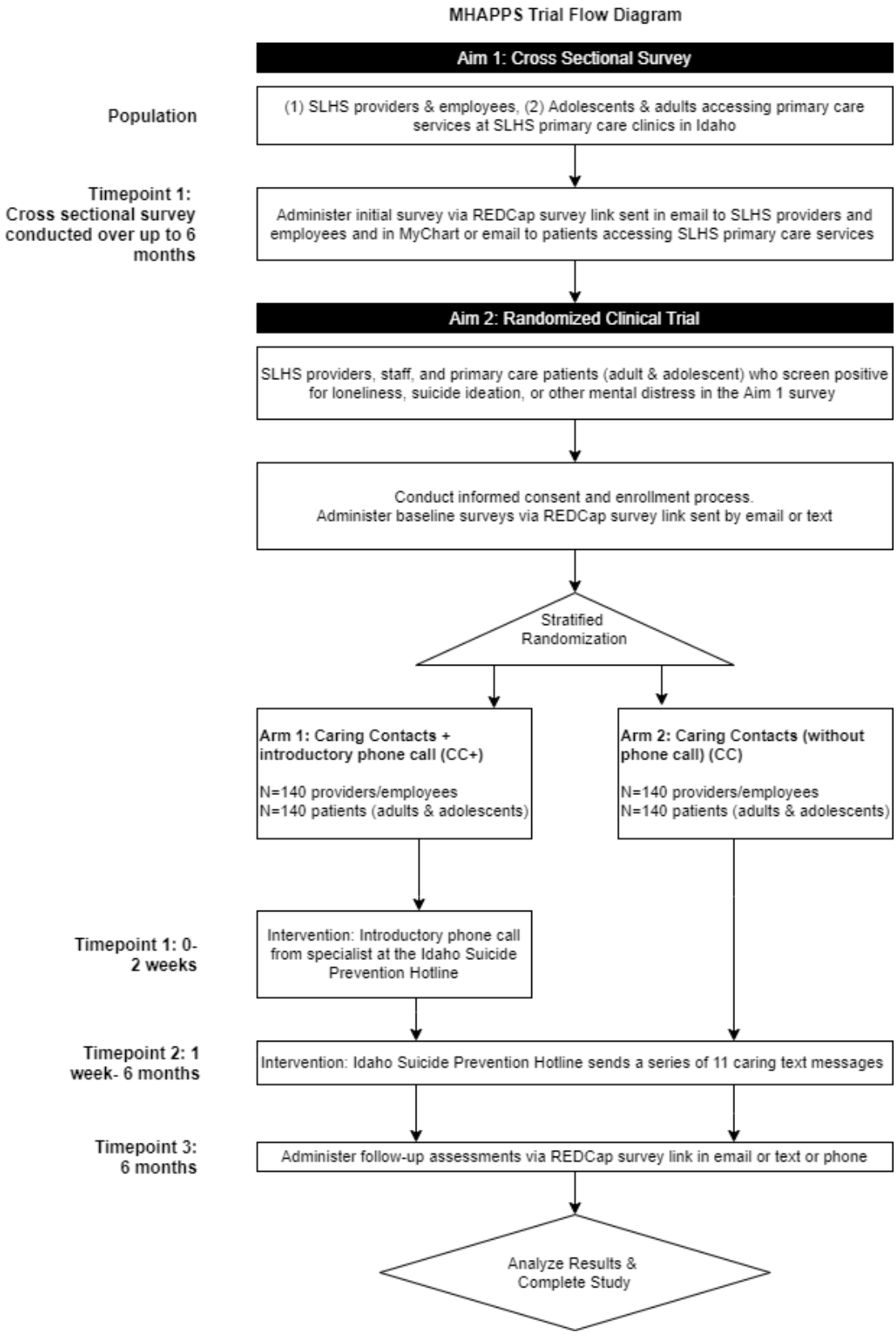
Aim 2: Compare the effectiveness of two versions of the Caring Contacts intervention ((1) introductory phone or video call (phone call) followed by text messages (CC+), versus (2) text messages alone (CC)) to reduce loneliness and improve mental health outcomes among SLHS patients, providers, and staff experiencing mental distress during the COVID-19 era.

3 DESIGN

Aim 1 will be addressed by a cross sectional survey of patients, providers, and staff to measure the prevalence of various measures of mental distress in the context of the COVID-19 pandemic.

Aim 2 will be investigated with a randomized controlled trial comparing the effectiveness of CC+ vs. CC. The study objectives, outcomes, and interventions are performed and assessed at the individual level. See Figure 1 for a flow diagram of the study.

Figure 1. MHAPPS Trial Flow Diagram



4 DATA SOURCE

All data for this study will be collected via online REDCap surveys. *Table 1* includes a list of all study variables, which aim they are associated with, and which tool will be used for assessment.

Table 1: MHAPPS Trial Variables & Other Data Elements

Variables			
Outcomes	**Tool**	**Aim Collected**	**Source**
Loneliness	NIH Toolbox – Social Relationship Scales - Loneliness Scale	Aim 1	REDCap
Suicidal ideation & behavior	C-SSRS Screener	Aim 1	REDCap
Recent suicide attempts, suicide completion, & self-harm; lethal means	6-Months Suicide Attempts Survey	Aim 1	REDCap
Perceived Burdensomeness; Thwarted Belongingness	INQ-15	Aim 1	REDCap
Stress	NIH Tool Box – Perceived Stress 10	Aim 1	REDCap
Alcohol and Illicit drug use	Youth Risk Behavior Survey Questions	Aim 2	REDCap
Depression	PHQ-9	Aim 1	REDCap
Anxiety	GAD-7	Aim 1	REDCap
Uptake of outpatient mental health treatment	n/a	Aim 2	REDCap
Exposure (Baseline)	**Tool**	**Aim Collected**	**Source**
Age in years, age category (adult/peds)	Baseline Survey	Aim 1	REDCap
Sex	Baseline Survey	Aim 1	REDCap
Race and ethnicity	Baseline Survey	Aim 1	REDCap

Address including Zip code of residence (urban/rural)	Baseline Survey	Aim 1	REDCap
County of residence (urban/rural)	Baseline Survey	Aim 1	REDCap
Gender identity & sexuality	Baseline Survey	Aim 2	REDCap
Religion	Baseline Survey	Aim 2	REDCap
Occupation	Baseline Survey	Aim 1	REDCap
Education	Baseline Survey	Aim 1	REDCap
Living Situation (permanent, temporary)	Baseline Survey	Aim 1	REDCap
Number of People at Home	Baseline Survey	Aim 1	REDCap
Perceived Risk for COVID-19	Baseline Survey	Aim 1	REDCap
Masking Practices	Baseline Survey	Aim 1	REDCap
Social Distancing Practices	Baseline Survey	Aim 1	REDCap
Domestic Violence	Baseline Survey	Aim 2	REDCap
Local COVID-19 Prevalence			CDC or Idaho Department of Health & Welfare Prevalence Data
Process Variables			**Source**
Aim 1 proportion of eligible participants completing survey			REDCap
Aim 2 proportion of eligible participants enrolled			REDCap
Participant feedback on phone call (CC+ arm only)			REDCap
Participant feedback on content and timing of caring messages			REDCap
"Dose" of follow-up contact: timing, type (phone vs text), and			Mosio

number of attempted and successful contacts from the Hotline		
"Dose" of Caring Contact text messages: date/time of each text message sent (* 11 text messages)		Mosio
Other Variables	**Tool**	**Source**
Does cell phone on record belong exclusively to study participant or is it shared?	Baseline Survey	REDCap
Alternative modes of contact	Contact Form	REDCap
Death		State Vital Records

4.1 Data Management

This study will employ a comprehensive data management plan. The results from the screening and baseline questionnaires (as well as informed consent documents) will be directly entered in REDCap. Participants will be assigned a unique identifier; Aim 2 participants will be randomly assigned to one of the two intervention arms during the informed consent process just prior to study enrollment. Follow-up specialists at the Hotline will record any attempt at contact or successful contact made. The statistics team will compile data for each participant from REDCap on a routine basis for reports and to build and maintain a complete dataset.

The University of Washington Institute of Translational Health Sciences (ITHS) hosts REDCap, a secure, HIPAA-compliant web application, which will be used for building and managing online surveys and databases for this research. ITHS provides REDCap support and an array of research data curation and storage support. Other databases (such as Access, Excel) may be used for study management purposes; all such data will be kept on secured, password-protected computers.

5 ANALYSIS SETS/POPULATIONS/SUBGROUPS

We plan to survey up to 4,000 participants to address Aim 1. The total sample size for the Aim 2 trial is 660 participants, selected as a subset of those who complete the Aim 1 survey. For each stratum (patients, providers/staff), a sample of 330 participants (165 in each arm) allows 80% power to detect a difference of 5 units in the primary outcome (loneliness) if a minimum of 70% of participants are retained through study completion.

5.1 Inclusion & Exclusion Criteria: Aim 1

Provider & Employee Inclusion Criteria

- Provider or Employee at St. Luke's Health System
- Adults ≥ 18 years of age
- Proficient in spoken and written English language

Provider & Employee Exclusion Criteria

- Individuals who are unable or unwilling to provide informed consent to participate
- Individuals who are study staff for this study or the SPARC Trial
- Individuals who are participants in the SPARC Trial

Patient Inclusion Criteria

- Patient at a SLHS ED or primary care study site
- Current MyChart account user
- Adults ≥18 years of age
- Minors 12-17 years of age
- Proficient in spoken and written English language

Patient Exclusion Criteria

- Individuals who are unable or unwilling to provide informed consent to participate.
- Individuals who are participants in the SPARC Trial

5.2 Inclusion & Exclusion Criteria: Aim 2

Provider & Employee Inclusion Criteria

- Moderate or high score for loneliness, suicide ideation, psychological stress, anxiety, or depression:
    - NIH Toolkit Loneliness raw score of 13 or greater; or
    - C-SSRS score of 3 or greater; or
    - NIH Toolkit Perceived Stress raw score of 31 or greater for adults; or
    - GAD7 score of 11 or greater; or
    - PHQ9 score of 10 or greater
- Access to a phone for the duration of the study with the ability to receive text messages and phone calls

Provider & Employee Exclusion Criteria

- Individuals who are unable or unwilling to provide informed consent to participate
- Individuals who are study staff for this study or the SPARC Trial

Patient Inclusion Criteria

- Moderate or high score for loneliness, suicide ideation, psychological stress, anxiety, or depression:
    - NIH Toolkit Loneliness raw score of 13 or greater for adults and 16 or greater for adolescents; or

    - C-SSRS score of 3 or greater; or
    - NIH Toolkit Perceived Stress raw score of 31 or greater for adults or greater than 33 for adolescents; or
    - GAD7 score of 11 or greater; or
    - PHQ9 score of 10 or greater
        - *Note: validated youth versions of the NIH Toolkit assessments (loneliness and perceived stress), and PHQ-A tools will be used for adolescents; the C-SSRS and GAD7 tools are validated for use with both adults and adolescents*
- Access to a phone for the duration of the study with the ability to receive text messages and phone calls

Patient Exclusion Criteria

- Patients who are unable or unwilling to provide informed consent/assent to participate (or whose legally authorized representative is unable or unwilling to provide consent in the case of adolescents). Examples may include but are not limited to patients who present with acute or chronic cognitive impairment that would preclude their ability to consent (i.e. acute psychosis, intoxication, or intellectual disability).

5.3 Analysis Populations

1. Aim 1: All participants who complete the survey will be included in the data tabulations.
2. Aim 2:
    a. **Intention to Treat (ITT) Population**: Data for all participants that complete study enrollment and the baseline survey will be included in this dataset. The primary and secondary endpoints will be evaluated in this group.
    b. **Safety Analysis Population**: All safety analysis will include all participants who completed study enrollment and the baseline survey (equivalent to the ITT population).
    c. **Per-Protocol Analysis Population**: The per-protocol analyses will include the subset of participants who were retained for the duration of the study and completed the assigned intervention as outlined in protocol section 9.
    d. **Additional Populations**: Additional populations may be used to complete sensitivity analyses, for example, where missing data have been imputed using different techniques.

5.4 Subgroups

We plan to generate different effect estimates for (1) providers and staff, and (2) patients and are statistically powered to do so. We also plan to produce separate effect estimates for the following subgroups: adult vs adolescent patients, types of providers, low baseline C-SSRS score vs moderate to high baseline C-SSRS score, Hispanic vs non-Hispanic, female vs male, cisgender vs transgender or gender-nonconforming, heterosexual vs. homosexual, bisexual, or other; and urban vs rural; however, this study has not been specifically powered to identify differing treatment effects in each of these subgroups.

6 ENDPOINTS AND COVARIATES

6.1 Aim 2: Primary endpoint

The primary outcome is loneliness, measured at 6 months as a score using the NIH Toolbox Social Relationship Scales loneliness measure. The instrument results in an uncorrected standard score (T-score) as well as an age- and gender-corrected score for adults and for adolescent ages 8-17; higher scores indicate more loneliness. Analyses will be based on the age- and gender-corrected scores.

6.2 Secondary endpoints

- Secondary outcomes include suicide ideation; suicidal ideation and behavior will be measured as score at 6 months using the Columbia Suicide Severity Rating Scale (C-SSRS). Scores range from 0-6. Risk factors for suicide will be measured by the Interpersonal Needs Questionnaire (INQ-15). This survey measures feelings of perceived burdensomeness and thwarted belongingness; scores range from 15 to 105. The changes from baseline thwarted belongingness and perceived burdensomeness scales will be analyzed separately.
- Depression will be measured as score at 6 months in the Patient Health Questionnaire-9 (PHQ-9); this questionnaire results in a score between 0-27, with higher scores indicating worse depression. A score of 5-9 is considered minimal depression, 10-14 is considered mild major, 15-19 is moderate major, and ≥20 is severe major. Analysis will be based on the continuous score, with categories examined in exploratory/sensitivity analysis.

6.3 Exploratory endpoints

- Suicide attempts and lethal means will be summarized and compared across treatment arms. Questions will include the number of attempts and aborted attempts and lethal means.
- Psychological stress will be measured using the NIH Toolbox and assessed as change from baseline. This instrument results in a continuous T-score outcome.

- Changes in alcohol and illicit drug use will be described. Participants will be asked if they are current, ever, or never user of a variety of substances (tobacco, alcohol, marijuana, other illicit drugs), as well as the amount and frequency of use for current users. Participants will be asked about whether or not the amount they use has changed. Changes from baseline to 6 months in the current/ever/never question will be summarized for each substance, as well as reporting the number and percentage of participants with any increase or any decrease.
- The GAD-7 scale will be used to measure anxiety. Participants respond to seven items and receive possible scores of 0-21.

- Attendance at mental healthcare appointments will be measured as a dichotomous variable (yes/no) at 6 months through self-report.

6.4 Safety Outcomes

This protocol considers completed suicide, suicide attempts, and inpatient admission in the context of highly suicidal study participants as expected events. These will be routinely tracked as key safety outcomes. The following safety outcomes will be assessed for each participant at 6 months and reviewed by the Data and Safety Monitoring Board twice annually to determine whether the rate of safety outcomes differs by intervention arm:

- Death by suicide
- Attempted suicide
- Interrupted or aborted suicide attempt
- Psychiatric hospitalization for anxiety, depression, and/or suicidal ideation
- Medical hospitalization related to self-harm or attempted suicide
- Medical hospitalization related to unintentional overdose or substance use disorder

6.5 Covariates

Potential effect modifiers will be assessed and include age, sex, gender identity, sexuality, race/ethnicity, religion, employment, urban/rural residence, drug/alcohol use, suicidal ideation at baseline, baseline depression score, baseline anxiety score, baseline stress score, baseline quality of life, measures of socioeconomic status and/or other variables related to the specified outcomes of interest.

7 STATISTICAL METHODOLOGY

7.1 Aim 1:

7.1.1 Descriptive statistics

Descriptive statistics will be used to summarize cross sectional survey results as well as baseline demographic and other exposure variables. This will include number and percentage for categorical variables; for continuous variables mean and standard deviation will be tabulated, with median, first quartile, and third quartile included in at least internal reports.

7.1.2 Association of COVID attitudes, beliefs, and practices with outcomes

Linear regression models or generalized linear regression models (GLM) with robust standard errors will be used to assess the association between the primary and secondary outcomes and COVID attitudes, beliefs, and practices, adjusting for baseline demographic information. Each outcome will be modeled separately.

7.1.3 Variations in outcomes across type of employee

To assess how different types of healthcare providers are coping with COVID related mental distress and changes, we will use linear models with robust standard error, testing for a difference in reported distress levels across provider type while adjusting for baseline demographics. Employee categories will include:

- provider (physician and advanced practice providers such as physician assistants and nurse practitioners);
- nursing;
- social support services;
- other patient-facing, including pharmacy, medical assistants, and technicians; and
- non-clinical, including officer work, educators, senior management, and other.

7.2 Aim2:

7.2.1 Primary analysis

The effects of the intervention will be modeled separately in patients and providers/staff. **Linear regression with robust standard errors will be used to determine whether the primary outcome differs between CC+ vs CC**, adjusting for the baseline score as a precision variable. Multiple imputation will be used to account for missing data (see 7.3).

7.2.2 Secondary analyses

All outcomes will be modeled separately in the patient and provider/staff populations. Linear regression will be used to estimate the difference in means for continuous outcomes; for binary outcomes, GLM models with an identity link will allow for the estimation of the difference in proportions. Robust/sandwich standard errors will be used to allow for departures of the observed standard errors from classic model assumptions.

7.2.3 Subgroups

The primary outcome will be summarized in subgroups defined by:

- Age category: older adult (50+) vs. adult (26-49) vs younger adult (18-25) vs adolescent patients (<18),
- types of providers (patient-facing providers [physicians and APP; pharmacy, PAS, lab, specialists, technicians, MAs;; nursing; social support services] vs. non-clinical [office work and educators; senior management; and other]),
- baseline suicide risk (low baseline C-SSRS score vs moderate to high baseline C-SSRS score),
- Hispanic vs non-Hispanic,
- female vs male sex at birth,
- cisgender vs transgender or gender-nonconforming,
- heterosexual vs. homosexual, bisexual, or other, and
- urban vs rural.

7.2.4 Safety analysis

The number and proportion of subject experiencing the listed safety events, as well as any other reported safety events, will be summarized by cohort and treatment arm.

7.3 HANDLING OF MISSING DATA

Hot deck imputation will be used to account for missing 6-month outcomes. The 6-month outcomes will be sampled from complete cases on the same treatment arm, baseline loneliness score tertile, and study population. This method avoids making parametric

assumptions about the outcome distribution and preserves the relationship between measures (Andridge and Little (2010)). Twenty complete data sets will be imputed and Rubin's rule will be used to pool the results (Rubin 10987).

7.4 SENSITIVITY ANALYSIS

If the proportion of missing 6-month outcomes is greater than 10%, we will assess the sensitivity of the primary results to a variety of possible missingness generating models (Cro et al. 2020). Sensitivity analyses will explore the possibilities that missing data is more likely to reflect poor outcomes or the missing data mechanism varies by treatment arm. A tipping point analysis will be included.

7.5 PROGRAMMING PLANS

All data cleaning and programming will be done in SAS, R or RStudio.

8 References

Andridge RR, Little RJA. A Review of Hot Deck Imputation for Survey Non-response. *International Statistical Review* (2010), **78**: 40-64.

Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Statistics in Medicine* (2020) **39**:2815-2842.

Rubin, D.B. 1987. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons.