Statistical analysis plan

Optimizing a Mobile Mindfulness Intervention for ICU Survivors (LIFT2)

NCT04038567

Statistical Analysis Plan (SAP)

Title	Optimizing a self-directed mobile mindfulness intervention for improving cardiorespiratory failure survivors' psychological distress (LIFT2)
CRU/Department/Division/Centre	Pulmonary
IRB Number	Pro00100252
Investigators:	
Lead Investigator	Christopher Cox
Multiple PIs	Terri Hough, Marc Moss, Christopher Cox
Co-investigators	Laura Porter, Jeff Greeson, Tina Gremore
Biostatistician(s)	John Gallis
Supervising Biostatistician	Maren Olsen
Original Creation Date	01/08/2020
Version Date	11/28/2023
Project Folder Location	
Project Goal(s)	Manuscript
Submission Deadline(s)	

Investigator Agreement	☒ All statistical analyses included in an abstract or manuscript should reflect the work of the biostatistician(s) listed on this SAP. No changes or additional analyses should be made to the results or findings without discussing with the project biostatistician(s). ☒ All biostatisticians on this SAP should be given sufficient time to review the full presentation, abstract, manuscript, or grant and be included as co-authors on any abstract or manuscript resulting from the analyses. ☒ I have reviewed the SAP and understand that any changes must be documented. *Acknowledged by:* Christopher Cox *Date:* February 6, 2020
Activity Log	January 08, 2020 – John moved the SAP to this newly formatted file. January 31, 2020 – Added table of statistical power; weaved in some of the text from the protocol paper

	into the statistical analysis section. Added parameterization of the model and example SAS code. July 20, 2022 – Added more information on sample size. Now computing power for a range of effect sizes given 240 participants with 20% dropout (i.e., 192 participants after dropout). August 5, 2022 – Adding section about minimization; adding notes about determining a threshold June 23, 2023 – Clarifying that decisions will be made both in relation to effect size and p-value June 29, 2023 – Updated data decisions regarding windows August 22, 2023 – Updates to note that when PHQ-9 is the outcome, we do not adjust for PHQ-9 stratified; similarly for PHQ-10. November 28, 2023 – Error fixed in power calculation simulation code. (Accidentally powered on T3 main effect due to confusion about naming of time points.)

Contents

1 Study Overview

As survival has improved for the 2 million people with cardiorespiratory failure managed annually in US intensive care units (ICUs), it has become apparent that these patients suffer from severe and persistent post-discharge symptoms of psychological distress including depression, anxiety, and post-traumatic stress disorder (PTSD). However, few targeted interventions exist that are relevant to patients' experiences and that accommodate their many physical, social, and financial barriers to personalized care.

Mindfulness is a practice of non-judgmental awareness that can alleviate distress by uncoupling emotional reactions and habitual behavior from unpleasant symptoms, thoughts, and emotions.[1,2] Standard mindfulness training, typically provided face to face in group settings, has proven efficacious in improving psychological distress in various patient populations.[3,4] However, in-person therapy is infeasible for the many ICU survivors with new disabilities, financial distress, and great distance from referral centers.[5]

Therefore, we developed a telephone-delivered mindfulness-based training program for ICU survivors called LIFT and found that it was associated with improved symptoms of depression, anxiety, and PTSD in an uncontrolled pilot evaluation.[6] Next, we adapted the LIFT program to a self-directed mobile mindfulness training app to overcome the inconvenience of scheduled weekly telephone sessions, and then compared it to telephone-based mindfulness and an education control in a pilot randomized controlled trial (LIFT1). This trial provided compelling evidence for the mobile mindfulness app's feasibility, adherence, retention, and clinically meaningful impact.[7] It also showed how the LIFT program could be improved further by better targeting a population likely to respond, improving the app's delivery of content, and automating more features to improve user engagement.

Having addressed these gaps, we are now initiating a 2 x 2 x 2 factorial clinical trial (LIFT2) that is conceptualized as the Optimization Phase within the multiphase optimization strategy (MOST) framework. Through this trial, we will optimize mobile mindfulness by assessing three intervention components and their interactions. At the conclusion of this study involving 240 cardiorespiratory failure survivors with high levels of psychological distress just after hospital discharge, we will deliver a mobile mindfulness system fully optimized for usability, efficiency, scalability, and clinical impact that will be off-the-shelf ready for a next-step definitive RCT (LIFT3)—and can serve as a model for distance-based mind and body interventions.

The study time points are T1 (baseline & randomization), T2 (1 month post-randomization), and T3 (3 months post-randomization).

1.1 Statistical Aims

- Determine the effect of each of the three intervention components on change in PHQ-9 from randomization (T1) to 1 month post-randomization (T2).
- Assemble a multicomponent intervention package including the comparator level of components with main effect sizes of ≥ 1.6 units or $p<0.05$. Otherwise, include the default level, unless interactions of ≥ 1.6 units (or $p<0.05$) are present that would cause us to include the comparator level.
- Use other qualitative and quantitative criteria to potentially reconsider decisions.

2 Study Population

2.1 Inclusion Criteria

See protocol.

2.2 Exclusion Criteria

See protocol.

2.3 Sample size

As described below, we need to have sufficient statistical power to detect main effects or interactions of at least 1.6 points on the PHQ-9—our decision-making benchmark. We calculated power for the test of a main effect or factor interaction in a constrained longitudinal model via simulation. Based on LIFT pilot data, the baseline standard deviation of PHQ-9 was assumed to be 5.3 and the covariance between time points ranged from 3.9 to 11.5 (**Table 1**). We examined a range of sample sizes at T2, from 120 participants up to 200 participants.

However, given our experience in a past psycho-behavioral intervention trial with nearly 40% dropout,[8] we revised our target sample size to 240 total participants. By targeting 240 randomized participants, we will have ample power to detect factor main and interaction effects even with deviations from our assumptions (e.g., higher dropout rate, misspecified covariance matrix, etc.). We simulated 500 different datasets under several smaller effect sizes of PHQ-9 (**Table 2**). If our assumptions about dropout are correct, then we have 80% power to detect a main effect or interaction effect of 1.6 on the PHQ-9 scale. Appendix B displays the Stata code for simulating power.

Table 1. Assumed covariance matrix used to generate simulated PHQ-9 data

	T1	T2	T3
T1	28.0		
T2	9.7	19.3	
T3	3.9	11.5	11.8

*Note: values on the diagonals are variances; based on LIFT pilot data

Table 2. Power by PHQ-9 effect size (mean differences), N=192 participants after 20% dropout (recruited N=240)

Regression coefficient	Effect size (mean difference)	Power
0.5	1.0	42.2%
0.55	1.1	47.8%
0.6	1.2	54.6%
0.65	1.3	61.2%
0.7	1.4	67.4%
0.75	1.5	75.4%
0.8	1.6	79.8%

*Note that when using effect coding, the effect size is 2 times the regression coefficient. All effect sizes are on the nominal (unstandardized) scale.

2.4 Randomization

The combination of the three factors each at two levels results in 8 experimental conditions. Allocation of the patients into one of the 8 experimental conditions (240 patients randomized with 30 in each condition) occurs at the time of T1 data collection completion (i.e., the first data collection after arrival home from the hospital). To ensure balance between the conditions on these important confounders, allocation is stratified by site (Duke, Colorado, Washington/Oregon), medical vs. surgical ICU vs. not applicable, depression symptom score at T1 (PHQ-9 < 15, ≥15), physical symptom score at T1 (PHQ-10 < 10, ≥10), and age (< 50, ≥50). Because of the large number of stratification variables relative to the experimental condition size, treatment assignment is conducted via a dynamic allocation minimization method.[9] This algorithm is programmed directly within the LIFT mobile app platform.

2.5 Data Acquisition

Fill in all relevant information:

Study design	Factorial trial
Data source/how the data were collected	Primary data collection from participants in North Carolina, Washington, and Colorado
Contact information for team member responsible for data collection/acquisition	Santos Bermejo santos.bermejo@duke.edu
Date or version (if downloaded, provide date)	TBD
Data transfer method and date	REDCap to Box
Where dataset is stored	A Duke Box folder called PROJECT_DGHI_LIFT

2.6 Data Decisions

Per the protocol, all participants with measures out of window will be removed from the analysis. The windows are as follows:

Time Point	**Window**
1 month post-randomization (T2)	Up to 60 days post-randomization (+30 days from 1 month
3 months post-randomization (T2)	Up to 120 days post-randomization (+120 days from 3 months)

Records outside of that window are identified in the final dataset using an indicator called “outofwindow”.

3 Outcomes, Exposures, and Additional Variables of Interest

3.1 Primary Outcome(s)

- The primary outcome is PHQ-9 at 1-month post-randomization (T2).

3.2 Secondary Outcome(s)

- PHQ-9 at 3-months post-randomization (T3).
- GAD-7 at 1 and 3 months post-randomization
- PHQ-10 at 1 and 3 months post randomization
- PTSS at 1 and 3 months post randomization
- Quality of life: visual analog scale at 1 and 3 months post randomization

3.3 Other outcomes

- Mindfulness Attention Awareness Scale (MAAS) at 1 and 3 months post randomization

3.4 Other variables of interest.

- Feasibility, acceptability, usability, and participant feedback gathered through semi-structured interviews.

3.5 Primary predictors

The primary predictors are the three intervention components listed in table 3. Detailed descriptions of these components are provided in the protocol paper.[10]

Table 3		
Intervention component	**Default Level**	**Comparator Level**
Dose	Low dose	High dose
Non-responders	App-based	Therapist-based
Initiation	App-based	Therapist-based

4 Statistical Analysis Plan

The main goal of these intention-to-treat analyses is to estimate main effects and interactions of combined experimental conditions to be able to determine which LIFT use case is best, i.e., should any of the factors should move from the "low" to the "high" levels (**Table 1**). Our overall analytic approach will be guided by frameworks presented by Collins and Collins & Kugler.[11,12] To understand general patterns in the data, we will first calculate raw means, medians, and measures of variability at each time point for the primary and secondary outcomes. These will be grouped by the main effect of each factor (e.g. standard vs high-dose), collapsed over the other factors. We will also examine these descriptive statistics by groups as defined by the two-way and three-way interactions of the factors.

The primary quantitative outcome is the PHQ-9. The primary time point of interest is 1-month post-randomization (T2). We will estimate the main effects of each intervention component on the primary outcome and of all pairwise interactions of those components at each follow-up time point using a general linear model with unstructured covariance to take into account repeated measures on individuals over time. To estimate effects of each intervention component on PHQ-9, the model will include as outcomes PHQ-9 score at baseline (T1) and each follow-up time point (T2, T3). The model will include fixed effects for: time point (T2 and T3), the time-by-component and time-by-pairwise up to five-way interactions, and the variables used in the minimization allocation algorithm. We do not include the effects of the components at

baseline in the statistical model. This constrains the baseline comparisons to be equal, which is appropriate in a randomized trial and increases power.[13,14] The model will allow us to estimate PHQ-9 change at the primary time point of interest (T2) and the final follow-up time point (T3). The constrained longitudinal model is parameterized below.

$$Y_{ij} = \beta_0 + \beta_1 c1_{ij} T_{ij} + \beta_2 c2_{ij} * T_{ij} + \beta_3 c3 * T_{ij} + \beta_4 c1c2 * T_{ij} + \beta_5 c1c3 * T_{ij} + \beta_6 c2c3 * T_{ij} + \beta_7 c1c2c3 * T_{ij} + \beta_8 * T_{ij} + \epsilon_{ij},$$
$$\epsilon_{ij} \sim N(0, \boldsymbol{\Sigma}_{ij}),$$

where $c1, c2, c3$ are the intervention components (effect coded) assigned to person i at time j ($i = 1, \ldots, n; j = 0,1,2$), Y_{ij} is the outcome for participant i at time j, T_{ij} is the indicator of person i at time point j (with T1 coded as 0). Note that in this model we assume time 0 is the reference for the T_{ij} variable, and hence the effects of the intervention components are assumed to be 0 at time 0. $\boldsymbol{\Sigma_{ij}}$ is the residual correlation matrix, which we will assume to be unstructured, to take into account correlation within individuals across time. Example SAS code for fitting this model is shown in **Appendix A**. Note also that in practice we will adjust for the variables upon which the allocation was minimized. However, when PHQ-9 is the outcome, we will not adjust for baseline PHQ-9, and similarly for PHQ-10.

After examining the model, we will assemble a multicomponent intervention package. In our original protocol paper, we powered based on a 2-unit main effect, but because of increase in power due to an increase in sample size, we lowered the threshold to 1.6 units. If a component has a main effect on PHQ-9 at T2 of ≥ 1.6 units or a p-value < 0.05, and no clinically significant interaction (i.e., achieving the benchmark) with another component, then the superior level of the component will be retained for the intervention package. Otherwise, if there is no clinically significant main effect or interaction (differences less than the benchmark of 1.6 units or p-value > 0.05), the default (reference) level of the component will be retained. We will reconsider inclusions based on the presence of large (effect ≥ 1.6 units or p-value < 0.05) interactions. This decision making is based on the approach outlined in Collins et al.[15]

Although the secondary time point (T3) will not be used in the *primary* decision-making process, we may reconsider our inclusions (i.e., combinations of high and low levels) if there is a large change in effect between T2 and T3. We will then examine secondary outcomes, as well as feasibility and acceptability metrics, as a next step in the decision-making process. We may reconsider our inclusions if there is a large effect in the opposite direction from the primary model, or if feasibility and acceptability metrics, for example, indicate that the "high" level of the component is not acceptable to participants.

Given the factorial design, we will use effect coding, rather than dummy coding, for analyzing the effects of the intervention components.[16] If the sample size is equal per condition, all of the tests of main effects and interactions are uncorrelated—that is, the main effect of a condition is the same even if other treatment conditions and interactions are included in the statistical model. Even with unequal sample sizes across conditions (as may occur with differential dropout by condition), if the imbalance is minor, the correlations between effects should be small.[17]

Missing outcome data due to dropout or missing intermediate visits is expected to be at most 20%. Because the model will be fit using a full maximum likelihood method, we will be able to account for predictors of missingness in the model in order to obtain valid estimates of the main component effects, thanks to the property that the response can be missing at random (MAR) as defined by Little and Rubin.[18] In practice, we will compare baseline characteristics of completers and non-completers. If we find that any covariates predict missingness, we will adjust for these variables in the model as a sensitivity analysis.

5 References

1. Kabat-Zinn J. Full catastrophe living. New York ,NY: Delta Publishing, 1990.
2. Ludwig DS, Kabat-Zinn J. Mindfulness in medicine. JAMA 2008;300(11):1350-2. (In eng). DOI: 10.1001/jama.300.11.1350.
3. Fjorback LO, Arendt M, Ornbol E, Fink P, Walach H. Mindfulness-based stress reduction and mindfulness-based cognitive therapy: a systematic review of randomized controlled trials. Acta Psychiatr Scand 2011;124(2):102-19. (In eng). DOI: 10.1111/j.1600-0447.2011.01704.x.
4. Gotink RA, Chu P, Busschbach JJ, Benson H, Fricchione GL, Hunink MG. Standardised mindfulness-based interventions in healthcare: an overview of systematic reviews and meta-analyses of RCTs. PLoS One 2015;10(4):e0124344. DOI: 10.1371/journal.pone.0124344.
5. Cox CE, Docherty SL, Brandon DH, et al. Surviving critical illness: acute respiratory distress syndrome as experienced by patients and their caregivers. Crit Care Med 2009;37(10):2702-8. (In eng). DOI: 10.1097/CCM.0b013e3181b6f64a.
6. Cox CE, Porter LS, Buck PJ, et al. Development and preliminary evaluation of a telephone-based mindfulness training intervention for survivors of critical illness. Annals of the American Thoracic Society 2014;11(2):173-81. DOI: 10.1513/AnnalsATS.201308-283OC.
7. Cox CE, Hough CL, Jones DM, et al. Effects of mindfulness training programmes delivered by a self-directed mobile app and by telephone compared with an education programme for survivors of critical illness: a pilot randomised clinical trial. Thorax 2019;74(1):33-42. DOI: 10.1136/thoraxjnl-2017-211264.
8. Cox CE, Hough CL, Carson SS, et al. Effects of a Telephone- and Web-based Coping Skills Training Program Compared with an Education Program for Survivors of Critical Illness and Their Family Members. A Randomized Clinical Trial. Am J Respir Crit Care Med 2018;197(1):66-78. DOI: 10.1164/rccm.201704-0720OC.
9. Han B, Enas NH, McEntegart DJSim. Randomization by minimization for unbalanced treatment allocation. 2009;28(27):3329-3346.
10. Cox CE, Olsen MK, Gallis JA, et al. Optimizing a self-directed mobile mindfulness intervention for improving cardiorespiratory failure survivors' psychological distress (LIFT2): Design and rationale of a randomized factorial experimental clinical trial. Contemp Clin Trials 2020:106119.
11. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST): Springer, 2018.
12. Collins LM, Kugler KC. Optimization of multicomponent behavioral, biobehavioral, and biomedical interventions: Advanced topics. New York, NY: Springer; 2018.
13. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis. 2nd ed. Hoboken, N.J.: Wiley, 2011.
14. Lu K. On Efficiency of Constrained Longitudinal Data Analysis Versus Longitudinal Analysis of Covariance. Biometrics 2010;66(3):891-896. (In English). DOI: http://dx.doi.org/10.1605/01.301-0012175889.2010.
15. Collins LM, Trail JB, Kugler KC, Baker TB, Piper ME, Mermelstein RJ. Evaluating individual intervention components: making decisions based on the results of a factorial screening experiment. Transl Behav Med 2014;4(3):238-251.
16. Kugler KC, Trail JB, Dziak JJ, Collins LM. Effect coding versus dummy coding in analysis of data from factorial experiments. The Methodology Center, The Pennsylvania State University 2012.
17. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial experiments: efficient tools for evaluation of intervention components. Am J Prev Med 2014;47(4):498-504.
18. Little RJA, Rubin DB. Statistical analysis with missing data. New York: Wiley, 1987.

6 Appendices

6.1 Appendix A

SAS code for fitting the general linear model with unstructured covariance and constrained so that baseline is equal.

```
PROC MIXED DATA=analysis METHOD=REML CL noclprint=10;
        title "cLDA model";
        CLASS recordid time mz_icu mz_age /*mz_phq9*/ mz_phq10 mz_site;
        model phq9total =

        t2 t3

        /* regression coefficients at 1-month post-intervention */
        t2*dose t2*symp t2*call t2*dose*symp t2*dose*call t2*symp*call t2*dose*symp*call

        /* regression coefficients at 3-month post-intervention */
        t3*dose t3*symp t3*call t3*dose*symp t3*dose*call t3*symp*call t3*dose*symp*call

        mz_icu mz_age /*mz_phq9*/ mz_phq10 mz_site / solution;


        /* 1-month post-intervention effects (2*regression coefficient) */
        * MAIN EFFECTS;
        estimate "dose main effect t2"        dose*t2 2;
        estimate "symptom main effect t2"     symp*t2 2;
        estimate "call main effect t2"        call*t2 2;

        * TWO-WAY INTERACTIONS;
        estimate "dose, symptom interaction effect t2"      dose*symp*t2 2;
        estimate "dose, call interaction effect t2"         dose*call*t2 2;
        estimate "symptom, call interaction effect t2"      symp*call*t2 2;

        * THREE-WAY INTERACTION;
        estimate "dose, symptom, call interaction effect t2" dose*symp*call*t2 2;

        /* 3-month post-intervention effects (2*regression coefficient) */
        * MAIN EFFECTS;
        estimate "dose main effect t3"        dose*t3 2;
        estimate "symptom main effect t3"     symp*t3 2;
        estimate "call main effect t3"        call*t3 2;

        * TWO-WAY INTERACTIONS;
        estimate "dose, symptom interaction effect t3"      dose*symp*t3 2;
        estimate "dose, call interaction effect t3"         dose*call*t3 2;
        estimate "symptom, call interaction effect t3"      symp*call*t3 2;

        * THREE-WAY INTERACTION;
        estimate "dose, symptom, call interaction effect t3" dose*symp*call*t3 2;

        REPEATED time / TYPE=un SUBJECT=recordid R RCORR;
RUN;
```

6.2 Appendix B

Stata code for simulation of power

```
/* Program to simulate data */
program powersimu, rclass
        version 16.0
        syntax, b1(real) [alpha(real 0.05)]

        clear
        set obs 192
        gen id = _n in 1/192

        *correlation matrix; estimates from PHQ-9 data from LIFT1 trial
        matrix V = (27.95, 9.67, 3.91 \ ///
                    9.67, 19.28, 11.54 \ ///
                    3.91, 11.54, 11.78)
        matrix list V
        matrix M = (0 \ 0 \ 0)

        *random effect drawn from a multivariate normal distribution
        drawnorm e1 e2 e3, means(M) cov(V)

        *check that the covariances line up with expected
        correlate e1-e3, covariance

        *set the 8 different intervention versions
        gen condition=1 in 1/24
        replace condition=2 in 25/48
        replace condition=3 in 49/72
        replace condition=4 in 73/96
        replace condition=5 in 97/120
        replace condition=6 in 121/144
        replace condition=7 in 145/168
        replace condition=8 in 169/192

        *set up the effect coded intervention components
        gen dose=1 if inlist(condition,1,2,3,4)
        replace dose=-1 if inlist(condition,5,6,7,8)

        gen symptoms = 1 if inlist(condition,1,2,5,6)
        replace symptoms = -1 if inlist(condition,3,4,7,8)

        gen call = 1 if inlist(condition,1,3,5,7)
        replace call = -1 if inlist(condition,2,4,6,8)

        *reshape to long format for fitting the statistical model
        reshape long e, i(id) j(time)

        *dummy code time
        replace time = time-1
        gen time1 = 1 if time == 1
        gen time2 = 1 if time == 2
        replace time1 = 0 if time1 != 1
        replace time2 = 0 if time2 != 1

        *rename variables for easier coding
        rename (dose symptoms call) (d s c)
        *note that t2 is the primary time point
        rename (time1 time2) (t2 t3)

        *generate interactions ahead of time
        gen ds = d*s
        gen dc = d*c
        gen sc = s*c
        gen dsc = d*s*c

        foreach x in t2 t3 {
                gen d`x' = d*`x'
                gen s`x' = s*`x'
                gen c`x' = c*`x'
                gen ds`x' = d*s*`x'
                gen dc`x' = d*c*`x'
                gen sc`x' = s*c*`x'
                gen dsc`x' = d*s*c*`x'
        }

        *generate PHQ-9 with a main effect on dose at t2. All other effects are null.
        gen phq9 = 13+`b1'*dt2+0*st2+0*ct2+0*dst2+0*dct2+0*sct2+0*dsct2+0*dt3+0*st3+0*ct3+0*dst3+ ///
```

```
                        0*dct3+0*sct3+0*dsct3+e

        *run mixed model with unstructured covariance on the residuals
        mixed phq9 t2 t3 dt2 st2 ct2 dst2 dct2 sct2 dsct2 dt3 st3 ct3 dst3 dct3 sct3 dsct3 ///
                || id:, noconst residuals(uns, t(time)) reml stddeviations

        *obtain the p-value on the interaction of dose and T2
        matrix a =r(table)
        local p1=a[4,3]
        return scalar pvalue=`p1'
        *determine whether less than alpha
        return scalar reject=(`p1'< `alpha')

end
```

/* Program to run the simulation multiple times */

```
capture program drop power_cmd_powersimu
program power_cmd_powersimu, rclass
version 16.0
   syntax, b1(real) power(real) ///
        [alpha(real 0.05) ///
        reps(integer 100) ]

   *call the powersimu program
   simulate reject=r(reject), reps(`reps') seed(12345):    ///
   powersimu, b1(`b1') alpha(`alpha')

   quietly summarize reject

   *return results
   return scalar power = r(mean)
   return scalar N = 192
   return scalar alpha = `alpha'
   return scalar b1=`b1'
end
```

/* Program to initialize parameters */

```
capture program drop power_cmd_powersimu_init
program power_cmd_powersimu_init, sclass
   sreturn clear
   sreturn local pss_numopts "b1"
   sreturn local pss_colnames "b1"
end
```

/* Run the "program to simulate data" multiple times */

```
power powersimu, reps(500) b1(0.5(0.05)0.80) power(.80) ///
  table graph(xdimension(b1) ydimension(power))

graph export "Power_Plot_PHQ-9_ES.emf", as(emf) replace
graph export "Power_Plot_PHQ-9_ES.pdf", as(pdf) replace
```