Appendix A. Supplemental Methods

Study Hypotheses

Hypothesis 1a: Compared to the waitlist control condition, stress symptoms will be relieved at 12 weeks for MBSR, the Daily Examen, and Stress Proofing when evaluated independently.

Hypothesis 1b: Compared to the waitlist control condition, stress symptoms will be relieved at 24 weeks for MBSR, the Daily Examen, and Stress Proofing when evaluated independently.

Hypothesis 2a: Compared to the waitlist control condition, anxiety symptoms will be relieved at 12 weeks for MBSR, the Daily Examen, and Stress Proofing when evaluated independently.

Hypothesis 2b: Compared to the waitlist control condition, anxiety symptoms will be relieved at 24 weeks for MBSR, the Daily Examen, and Stress Proofing when evaluated independently.

Hypothesis 3a: Compared to the waitlist control condition, depression symptoms will be relieved at 12 weeks for MBSR, the Daily Examen, and Stress Proofing when evaluated independently (exploratory outcome).

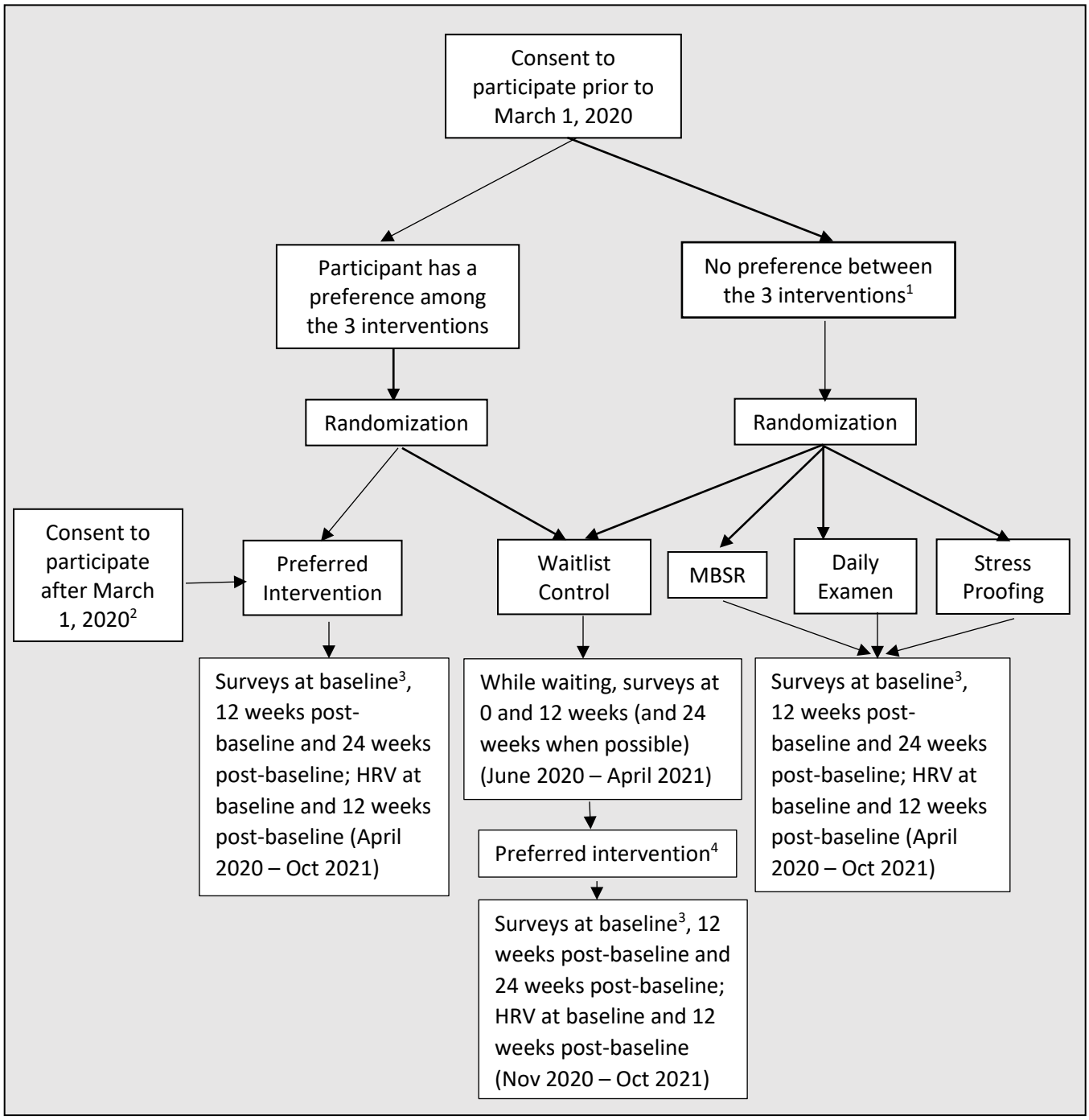
Hypothesis 3b: Compared to the waitlist control condition, depression symptoms will be relieved at 24 weeks for MBSR, the Daily Examen, and Stress Proofing when evaluated independently (exploratory outcome).

Hypothesis 4: Compared to the waitlist control condition, HRV will be improved at 12 weeks for MBSR, the Daily Examen, and Stress Proofing when evaluated independently.

Hypotheses 5a-5d: Participants who had a stated preference and received that intervention (i.e. MBSR, the Daily Examen, and Stress Proofing combined) will experience larger between-arm (waitlist vs non-waitlist) differences in improvements on a) stress symptoms at 12 weeks; b) anxiety symptoms at 12 weeks; c) depression symptoms at 12 weeks; and d) HRV at 12 weeks, when compared to no-preference participants randomly assigned across interventions and waitlist (exploratory outcome).

Hypotheses 5e-5g: Participants who had a stated preference and received that intervention (i.e. MBSR, the Daily Examen, and Stress Proofing combined) will experience larger between-arm (waitlist vs non-waitlist) differences in improvements on e) stress symptoms at 24 weeks, f) anxiety symptoms at 24 weeks, and g) depression symptoms at 24 weeks, when compared to no-preference participants randomly assigned across interventions and waitlist (exploratory outcome).

Supplemental Methods Figure A1. Pandemic-adapted Selah study design: A partially-randomized waitlist-controlled preference trial design



[1] Participants who prefer 2 interventions equally and over the third will be first randomly assigned between the 2 preferred interventions, and then randomized into either their preferred intervention or waitlist control group.

[2] All participants who consented after March 1, 2020 were non-randomly assigned to their preferred intervention. Participants with no preference selected a workshop with dates of their choosing.

[3] Baseline indicates immediately before intervention start.
[4] After waitlist control participants finish giving control data, they may participate in the intervention they originally indicated as their preference or change it to reflect their current preference.

Supplemental Methods Table A1. Study assignment and allocation approach for each preference and enrolment date scenario

Preference Scenario	Assignment Approach
Enrolled prior to March 1, 2020	
No preference between interventions	Randomly assigned to one of the four study arms: three interventions without waitlist and the waitlist arm, with a 1:1:1:1 ratio
Preferred two interventions equally and over the third intervention	Randomly assigned to one of the two interventions with a 1:1 ratio, and then a fraction was randomly assigned to the waitlist arm, with a 3:1 non-waitlist vs waitlist ratio for MBSR and Stress Proofing and a 5:4 non-waitlist vs waitlist ratio for the Daily Examen (DE was preferred by more participants and this allowed the same number of participants to be randomized into each study arm)
Preferred one intervention among the three	Assigned to their preferred intervention and combined with participants with two top preferences who had been randomized to that intervention, and then randomly assigned to non-waitlist vs waitlist arms, with a 3:1 non-waitlist vs waitlist ratio for MBSR and Stress Proofing and a 5:4 non-waitlist vs waitlist ratio for the Daily Examen
Any of the above scenarios and are part of a married (or cohabitating) couple who both meet study criteria and enrolled	To avoid spillover effects, each couple was treated as if they were one person, i.e., assigning both spouses to the same intervention and randomizing them into a non-waitlist vs waitlist arm. When a couple had different preferences, one preference was randomly chosen as the couple's preference.
Seven clergy with an established meeting group (a covenant group)	The seven clergy jointly chose a single preferred intervention and were randomized together to the non-waitlist vs waitlist arm.
Enrolled after March 1, 2020	
All enrollees after March 1, 2020 with no preference, two equal preferences, or one preference, and whether a clergy couple or not	Participants answered treatment preference survey items but regardless of their answers, self-selected the intervention with intervention dates they most wanted from the full list of workshop options. None were randomized to the non-waitlist vs waitlist arms; they all were assigned to non-waitlist.

Heart Rate Variability (HRV) data collection procedures and processing

During the enrollment process, participants were asked survey questions to determine their eligibility for inclusion in HRV data collection. Participants were excluded from HRV data collection if they had underlying medical conditions, including a diagnosis of tachycardia; being pregnant or becoming pregnant during the course of data collection; being diagnosed with COVID-19; having a pacemaker; and documentation of other cardiovascular-related chronic or acute morbidities that could impact the integrity of HRV data (Supplemental Table A4).

Two weeks prior to the intervention, participants were mailed a box with a Bittium eMotion Faros 180 recording device with electrodes. Participants were asked to attend an online, synchronous study orientation one week prior to the start of their workshop. During the study orientation, participants were oriented to a video and brochure that we created to convey the instructions for HRV data collection (see https://spiritedlife.org/hrv/). Participants were taught to connect the heart rate recording device's two electrode leads to two pre-gelled (Ag/AgCl) disposable Ambu BlueSensor wet-gel ECG electrodes placed beneath the right clavicle and left ribcage. Participants were instructed to wear this ambulatory heart rate monitoring device for a 48-hour period during week 0 and week 12, during which time participants proceeded with their usual work, exercise, bathing, and sleep routines.

Heart rate was measured using continuous electrocardiographic (ECG) recording sampled at a rate of 1,000 Hz and used to calculate heart rate variability. Study staff imported the 48-hour ECG recording to Kubios HRV Premium V3.4.1 software [1], partitioned it into 5-minute segments, visually inspected it to allow for manual correction of ectopic beats, detrended it, and then subjected it to Kubios' automatic artefact correction algorithm [2]. Heart rate variability was indexed using the time-domain metric Root Mean Square of Successive RR Differences (RMSSD) because it is less affected by breathing and a more suitable outcome measure in ambulatory studies than frequency-domain measures [3]. Five-minute segments across 24 hours of recording were subject to a cosinor analysis using the Cosinor package for R, based on recommendations for the detection of circadian rhythmicity [4].

Additional details on survey measures

Demographics measures

Survey items were included to measure ***sex*** (male/female); ***age*** (in years); ***race; ethnicity;*** self-reported ***physical, mental, and behavioral health conditions***; ***marital status;*** and having ***children living at home***. To capture work-related characteristics that may relate to stress, we measured ***appointment effort*** (i.e. full-time vs part-time appointed at UMC), ***bi-vocational status*** (i.e. having a job in addition to serving as clergy), kind of ***clergy appointment*** (i.e., serving a church vs in another capacity), ***number of congregations the study participant was appointed to***, and ***number of congregants pastored by the participant***.

Clinically relevant measures

Physical activity levels were measured using the ***Godin-Shephard Leisure-Time Physical Activity Questionnaire*** [5], a self-report measure of how often one has engaged in physical activity, measured separately for strenuous, moderate, and mild exercise, in the past seven days and for how many minutes per time. We used self-reported weight and height to assess ***body mass index*** [6]. We used single items to assess average daily ***caffeine intake***, and average weekly ***alcohol consumption***.

Stress-related measures

In addition to the Calgary-Symptoms of Stress Inventory [7], the survey asked about self-reported ***financial stress*** (How stressful is your current financial situation for you? Not at all to extremely), ***number of hours worked*** per week, and ***overall life stress level*** at study registration.

Preference measures

We included on the baseline survey an item for ***preference for online vs in-person intervention***. At the time of enrollment, we included the Treatment Acceptability and Preferences Scale [8] for each intervention. We also measured at study registration whether participants had already been ***practicing the Daily Examen or MBSR***, separately, at least 3 times each week.

The protocol paper offers details on the timing of each measure [9].

Sample size

Estimates of baseline outcome levels and expected effect sizes for C-SOSI used data from our non-randomized pilot study conducted prior to the full trial [10]. The average baseline C-SOSI score was 0.92 (SD=0.46) across all interventions, with 12-week follow-up scores of 0.7 (SD=0.58) for MBSR, 0.55 (SD=0.36) for Stress Proofing, and 0.51 (SD=0.38) for the Daily Examen.

Given an alpha of 0.0167 (based on a Bonferroni correction to hypothesis tests for the effects of three interventions on C-SOSI scores), a per-arm sample size of 40 for Daily Examen, 47 for Stress Proofing, and 195 for MBSR (which had a larger standard deviation, resulting in a larger sample size) yielded 80% power to detect a between-arm difference in means at 12 weeks of 0.22 for MBSR, 0.37 for Stress Proofing, and 0.41 for Daily Examen for a two-sample t-test with unequal variances, allowing for loss-to-follow-up of 20% and a design effect of 1.3 (corresponding to an ICC of 0.027 and average cluster size of 12) to account for clustering caused by the group-based intervention delivery. We calculated the design effect this way: $Design\ Effect = 1 + \delta(n - 1)$ where δ is ICC and n is average cluster size. Note that this conservatively assumes there is clustering throughout the sample, however, we expect only partial clustering due to group treatment delivery.

Previous literature recommends defining a medium effect size for HRV as a standardized mean difference of 0.50 [9] A per-arm sample size of 140 was calculated to yield 80% power to detect an effect size of 0.50 for a two-sample t-test with an alpha of 0.0167, allowing for loss-to-follow-up of 20% and a design effect of 1.3 to account for group-based intervention delivery. We recognized that this sample size was ambitious and analysis of HRV data may lack adequate statistical power.

While preliminary published sample size calculation adjustment for multiple comparisons treated hypotheses of positive effects for each of the 3 interventions as disjunction testing (i.e. 3 tests), upon more careful consideration the study team determined that tests for the individual interventions should be considered individual testing and testing of two separate primary outcomes (C-SOSI and HRV) should be considered disjunction testing, thus alpha adjustment occurred separately within each of the 3 interventions and adjusted for two hypotheses based on the two primary study outcomes [12].

Propensity score methods to balance baseline characteristics between arms

As noted in the manuscript, use of the partially randomized preference design during the trial period meant that by design the analytic data would be a mix of randomized data (for trial participants that had no preference) and observational data (for trial participants that had a preference and were allowed to

select their intervention), which made it likely that treatment arms would be imbalanced on baseline characteristics in an unadjusted analysis. In addition, randomization was performed prior to baseline data collection, with substantial study dropout in the interim. Thus, statistical analysis necessitated incorporation of observational methods to rebalance the intervention arms on characteristics that may have influenced selection into a particular intervention. A propensity score covariate adjustment method [13] was selected using covariate balancing propensity scores [14] with multinomial specification in order to generate the probability of receiving immediate MBSR, Daily Examen, or Stress Proofing interventions, or being designated as a waitlist participant. Propensity score models were generated separately for use in C-SOSI and GAD-7 outcomes vs the PHQ-8 outcome vs the HRV outcomes so that baseline outcome levels could be used in the prediction of treatment receipt (the PHQ-8 outcome is missing in more participants than the C-SOSI and GAD-7 outcomes; the HRV outcomes are available only on a subset of participants). Separate models were also produced for trial vs. observational data. Analyses were performed using an as-treated estimand to reflect the combination of observational with randomized data.

Distributions of propensity scores for participants receiving each of the immediate interventions as well as the waitlist control were visualized using histograms. Covariate balance pre/post propensity score regression adjustment was assessed using regressions with indicators for intervention and covariate adjustment for the propensity score [15].

Propensity score methods for sensitivity analyses including the observational data

Propensity score and outcome regression specifications were run for the combined cohort and remained largely the same with the exception of an addition of a binary indicator in the propensity score models to flag participants that were part of the fully observational data vs those that provided trial period data. Post-waitlist baseline and pre-intervention baseline covariate values were combined with the original trial period baseline data to generate the propensity scores, participants that provided waitlist data and intervention period data had two baseline observations that contributed to the computation of the propensity scores.

Supplemental Methods Table A3. Baseline variables included in propensity score model, by analytic sample

Baseline Variable	Trial Phase			Pooled Trial and Post-Trial Phase		
	C-SOSI and GAD-7	PHQ-8	HRV	C-SOSI and GAD-7	PHQ-8	HRV
Age (range 26-80)	X	X	X	X	X	X
Sex (Male/Female)	X	X	X	X	X	X
Single-racial, non-Hispanic White (Yes/No)	X	X	X	X	X	X
Married or cohabitating with partner (Yes/No)	X	X	X	X	X	X
Any children at home (Yes/No)	X	X	X	X	X	X
Full-time clergy in UMC (Yes/No)	X	X	X	X	X	X
Bi-vocational (Yes/No)	X	X	X	X	X	X
Number of congregations appointed to (None/1/2+)	X	X	X	X	X	X
Number of congregants pastored (None/1-149/150+)	X	X	X	X	X	X
Hours per week worked as UMC clergy (range 0-80)	X	X	X	X	X	X
Alcoholic drink intake (range 0-5)	X	X	X	X	X	X
Caffeinated beverage intake (range 0-4)	X	X	X	X	X	X
Metabolic equivalents (METs) per week (range 0-476)	X	X	X	X	X	X
Practicing Daily Examen 3+ times a week at registration (Yes/No)	X	X	X	X	X	X
Practicing Mindfulness 3+ times a week at registration (Yes/No or missing)	X	X	X	X	X	X
Number of top preferences among Selah interventions (0/1/2)	X	X	X	X	X	X
Body Mass Index (range 18.1-55.8)	X	X	X	X	X	X
Self-endorsed high cholesterol (Yes, current or history/Never or missing)	X	X	X	X	X	X
Overall life stress at registration (range 0-4)	X	X	X	X	X	X
C-SOSI stress symptoms (range 0.04-3.27)	X		X	X		X
Self-endorsed anxiety (Yes, current or history/Never or missing)	X	X	X	X	X	X
GAD-7 anxiety symptoms (range 0-20)	X		X	X		X
Self-endorsed depression (Yes, current or history/Never or missing)	X	X	X	X	X	X
PHQ-8 depression symptoms (range 0-21)		X	X		X	X
HRV MESOR (range 7.2-128.7)			X			X
HRV Amplitude (range 0.4-70.5)			X			X
Late registrants (Yes/No)				X	X	X
In post-trial sample (Yes/No)				X	X	X

Model Building for Final Outcome Regression Models

Invitations for the waitlist surveys (in groups of approximately 20 to be in comparable in size to workshop groups) were agnostic to intervention preference or assignment, which raised concerns about the potential for time confounding. Thus, regressions were also adjusted for number of months from the start of overall survey data collection (April 2020) to each respective survey. We explored the functional form of time from options of linear, square, and cubic, using baseline levels of C-SOSI and GAD-7 scores (to ensure the interventions did not influence time trends), selecting the functional form with the best fit using Akaike Information Criterion (AIC) [16], confirming the plausibility of functional forms using lowess plots.

We additionally explored the possibility that random slopes for study time (weeks from baseline) may better account for within –person correlation over time, however, comparisons of AIC indicated that random intercepts alone performed as well or better than models that included individual level random slopes for study time.

Visualization of residuals was used to assess normality to confirm that the assumptions of the linear model were adequately met.

Details of Missing Data Methodology

Our base sample was composed of participants providing any survey or HRV data to the study. Missing data could arise via missing outcome data at any time point (baseline, 12-weeks, or 24-weeks) or missing baseline covariate data (needed for propensity score generation). With missing data greater than 5% on both outcome and covariates and missing completely at random (MCAR) not a plausible (e.g. participants with more stress are plausibly more likely to withdraw from the study) and missing not at random (MNAR) also not a plausible assumption given the richness of available data, multiple imputation using chained equations (MICE) was used [17], [18], [19] to produce an alternative set of estimates as a sensitivity analysis to be compared to the original complete case estimates. The imputation process was integrated in the analytic process using the following steps:

1. Imputation using MICE for all variable included in either the final regression model OR the propensity score model OR that could help predict missing values or the presence of missing data (see Table A2). Perform augmented regression in the presence of perfect prediction for categorical variables. Create 10 imputed datasets
2. Calculate propensity scores separately for each of the 10 imputation datasets as well as the original
3. Perform final regression analyses separately by imputation dataset and combine using Rubin's rules [20].

In trials with clustered data, ideally the imputation process would incorporate the clustered structure of the data into the imputation process in order to avoid increased risk of Type I error [21]. However, software limitations, multiple treatment groups, and low sample size made incorporation of clustering infeasible. Therefore, the central goal of the imputation process is to ascertain whether there was potential bias in the magnitude of the treatment effect, not to ascertain statistical significance.

Supplemental Methods Table A4. Variables included and types of regression used for multiple imputation using chained equations (MICE)

Variables included	Regression type	C-SOSI and GAD-7 based analytic sample	PHQ-8 based analytic sample	HRV based analytic sample
C-SOSI at baseline (range 0.04-3.27)	N/A - independent only	X		X
C-SOSI at 3 months follow-up (range 0-2.58)	Linear	X		X
C-SOSI at 6 months follow-up (range 0-2.44)	Linear	X		
GAD-7 at baseline (range 0-20)	Linear	X		X
GAD-7 at 3 months follow-up (range 0-17)	Linear	X		X
GAD-7 at 6 months follow-up (range 0-15)	Linear	X		
GAD-7 in 2019 from the Panel study (range 0-20)	Linear	X		
GAD-7 in 2021 from the Panel study (range 0-21)	Linear	X		
PHQ-8 at baseline (range 0-21)	Linear		X	X
PHQ-8 at 3 months follow-up (range 0-17)	Linear		X	X
PHQ-8 at 6 months follow-up (range 0-16)	Linear		X	
PHQ-8 in 2019 from the Panel study (range 0-18)	Linear		X	
PHQ-8 in 2021 from the Panel study (range 0-20)	Linear		X	
Metabolic equivalents (METs) per week at baseline (range 0-476)	Linear	X	X	X
Hours per week worked as UMC clergy at baseline (range 0-80)	Linear	X	X	X
Hours per week worked as UMC clergy in 2019 from the Panel study (range 10-80)	Linear	X	X	
Hours per week worked as UMC clergy in 2021 from the Panel study (range 0-80)	Linear	X	X	
Number of weeks from baseline to 3 months follow-up (range 7-19)	Linear	X	X	X
Number of weeks from baseline to 6 months follow-up (range 8-30)	Linear	X	X	

Number of months from Selah month 1 on calendar to baseline (range 0-12)	N/A - independent only	X	X	X
Number of months from Selah month 1 on calendar to 3 months follow-up (range 3-15)	Linear	X	X	X
Number of months from Selah month 1 on calendar to 6 months follow-up (range 6-15)	Linear	X	X	
Any children at home at baseline (Yes/No)	Logistic	X	X	X
Any children at home in 2019 from the Panel study (Yes/No)	Logistic	X	X	
Alcoholic drink intake at baseline (range 0-5)	Ordered logistic	X	X	X
Caffeinated beverage intake at baseline (range 0-4)	Ordered logistic	X	X	X
Number of congregants pastored at baseline (None/1-149/150+)	Ordered logistic	X	X	X
Number of congregants pastored in 2019 from the Panel study (range 18-1,400)	Linear	X	X	
Number of congregants pastored in 2021 from the Panel study (range 0-1,400)	Linear	X	X	
HRV MESOR at baseline (range 7.2-128.7)	Linear			X
HRV MESOR at 3 months follow-up (range 5.7-105.2)	Linear			X
HRV amplitude at baseline (range 0.4-70.5)	Linear			X
HRV amplitude at 3 months follow-up (range 0.4-82.1)	Linear			X
Overall life stress at registration (range 0-4)	N/A - independent only	X	X	X
Self-endorsed depression at baseline (Yes, current or history/Never or missing)	N/A - independent only	X	X	X
BMI at baseline (range 18.1-55.8)	N/A - independent only	X	X	X
Age at baseline (range 26-80)	N/A - independent only	X	X	X

Female	N/A - independent only	X	X	X
As-treated intervention arm	N/A - independent only	X	X	X

Note: these multiple imputations were done in the Trial sample and not the pooled sample

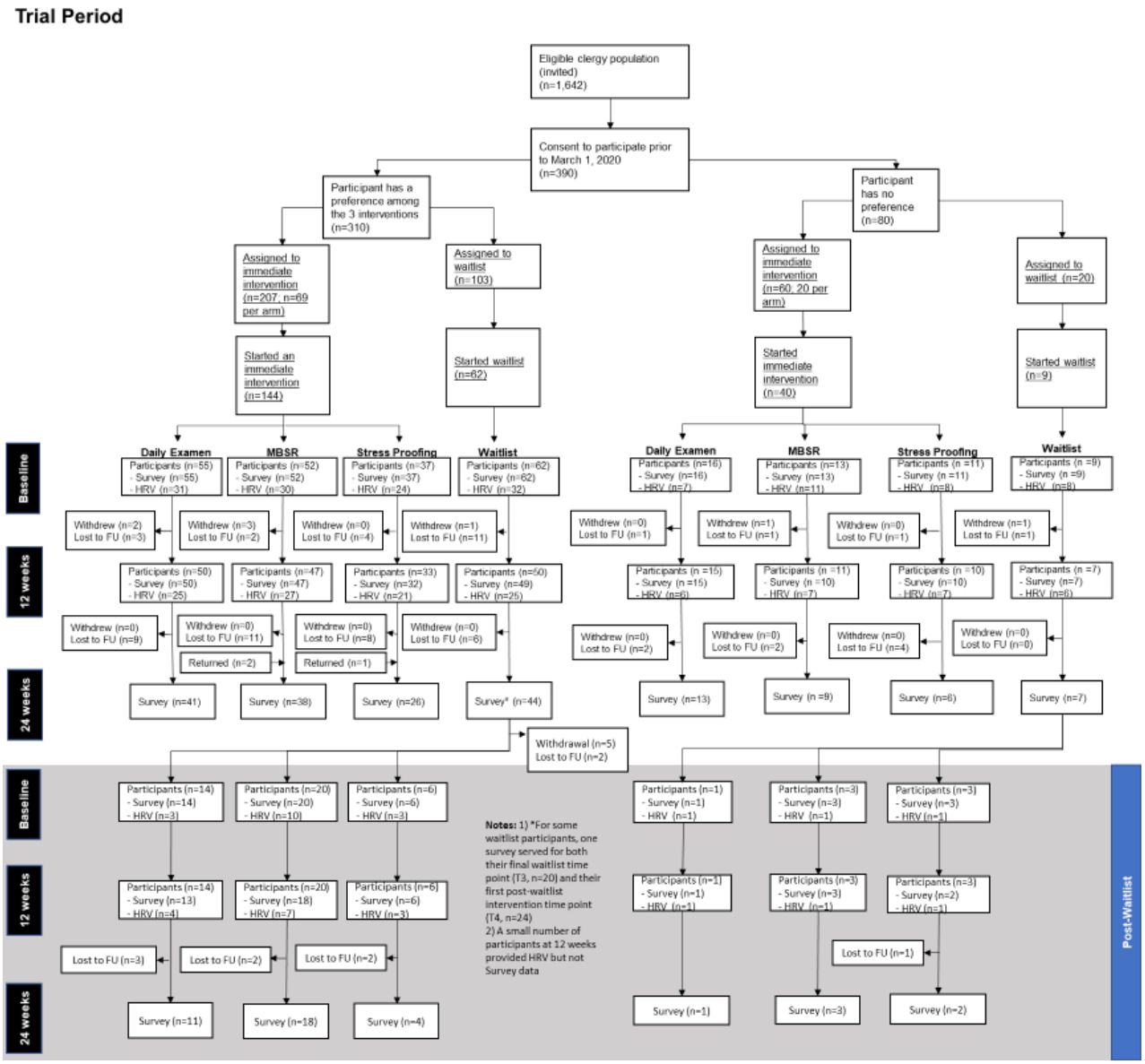
References

1. Tarvainen MP, Niskanen J-P, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV--heart ate variability analysis software. Comput Methods Programs Biomed.
2. Tarvainen M, Lipponen J, Niskanen J, Ranta-aho P. Kubios HRV Software Users Guide [Internet]. 2020 [cited 2021 May 18]. Available from: https://www.kubios.com/downloads/Kubios_HRV_Users_Guide.pdf.
3. Penttilä J, Helminen A, Jartti T, Kuusela T, Huikuri HV, Tulppo MP, et al. Time domain, geometrical and frequency domain analysis of cardiac vagal outflow: effects of various respiratory patterns. Clin Physiol. 2001 May;21(3):365–76.
4. Refinetti R, Lissen GC, Halberg F. Procedures for numerical analysis of circadian rhythms. Biol Rhythm Res [Internet]. 2007 [cited 2021 May 18];38(4):275–325. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3663600/
5. Godin G. The Godin-Shephard Leisure-Time Physical Activity Questionnaire. Health Fit J Can [Internet]. 2011 [cited 2021 May 18];4(1):18–22. Available from: https://hfjc.library.ubc.ca/index.php/HFJC/article/view/82
6. National Heart, Lung, and Blood Institute. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults. Bethesda (MD): National Heart, Lung, and Blood Institute; 1998.
7. Carlson LE, Thomas BC. Development of the Calgary Symptoms of Stress Inventory (C-SOSI). Int J Behav Med. 2007;14(4):249–56.
8. Sidani S, Epstein DR, Bootzin RR, Moritz P, & Miranda, J. (2009). Assessment of preferences for treatment: validation of a measure. Research in nursing & health, 32(4), 419-431.
9. Tice, L. C., Eagle, D. E., Rash, J. A., Larkins, J. S., Labrecque, S. M., Platt, A., Yao, J., & Proeschold-Bell, R. J. (2021). Rationale and preferences trial design to test three approaches to reduce stress symptoms among clergy. *Trials, 22,* 892. doi: 10.1186/s13063-021-05845-x
10. Proeschold-Bell, R. J., Eagle, D. E., Tice, L. C., Yao, J., Rash, J. A., Choi, J., Stringfield, B., & Labrecque, S. M. (2023). The Selah pilot study of spiritual, mindfulness, and stress inoculation practices on stress-related outcomes: A non-randomized participant preference control study. *Journal of Religion and Health, 62*(4), 2686-2710. doi: 10.1007/s10943-023-01848-x
11. Laborde S, Mosley, E, & Thayer, JF. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*, 8, 213.
12. Rubin M. When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. Synthese. 2021

13. Vansteelandt, S, & Daniel, RM. (2014), On regression adjustment for the propensity score, *Statistics in Medicine*, 33, pages 4053– 4072, doi: 10.1002/sim.6207
14. Imai K, Ratkovic M. Covariate balancing propensity score. J Royal Statistical Soc B. 2014;76:243–63.
15. Spreeuwenberg MD, Bartak A, Croon MA, Hagenaars JA, Busschbach JJV, Andrea H, et al. The Multiple propensity score as control for bias in the comparison of more than two treatment arms: An introduction from a case study in mental health on JSTOR. Med Care. 2010;48:166–74.
16. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr. 1974;19:716–23.
17. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol. 2017;17:162.
18. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. 2007;16(3):219-242. doi:10.1177/0962280206074463
19. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. Am J Epidemiol. 2010;171(5):624-632. doi:10.1093/aje/kwp425
20. Rubin DB. Multiple Imputation for nonresponse in surveys. Vol 81. Subsequent. Hoboken, N.J: Wiley-interscience; 2004:258.
21. Taljaard M, Donner A, Klar N. Imputation strategies for missing continuous outcomes in cluster randomized trials. Biom J. 2008;50:329–45.

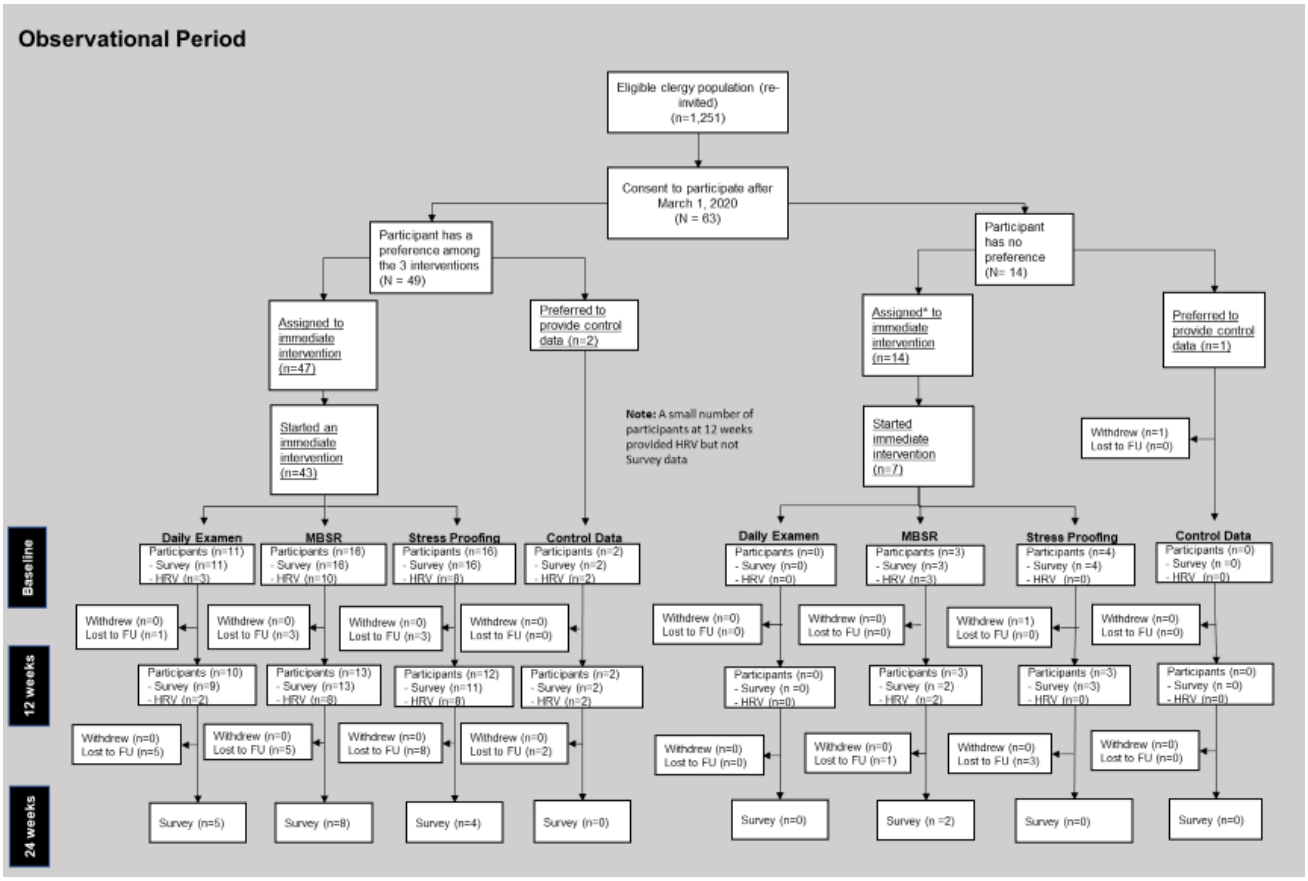
Appendix B. Supplemental Results

Supplemental Figure B1. Flow Chart for Trial Participants During the Trial Period and Post-waitlist Period



Note: Post-waitlist period data are analyzed as observational data.

Supplemental Figure B2. Flow Chart for Observational Participants



Supplemental Table B1. Baseline characteristics of Selah trial phase participants compared to 2019 Clergy Health Panel Study participants[1]

	Selah total (N = 255)	2019 Panel Study Population (N = 1278)	*p*-value[2]
Age (in years)			0.387
Mean (SD)	53.9 (11.2)	53.2 (12.3)	
Sex, n (%)			<0.001
Female	121 (47.5%)	429 (33.7%)	
Male	134 (52.5%)	843 (66.3%)	
Race & ethnicity (in mutual exclusive categories), n (%)			0.892
White and not Latinx	231 (90.6%)	1091 (88.8%)	
African American and not Latinx	15 (5.9%)	72 (5.9%)	
Asian American/Pacific Islander and not Latinx	2 (0.8%)	17 (1.4%)	
Native American and not Latinx	2 (0.8%)	13 (1.1%)	
Latinx	2 (0.8%)	18 (1.5%)	
Other, including bi-/multi-racial	3 (1.2%)	18 (1.5%)	
Marital & habitation status, n (%)			0.183
Not married, or married but separated/divorcing	27 (10.6%)	174 (13.7%)	
Married or cohabitating	228 (89.4%)	1098 (86.3%)	
Any children living at home, n (%)			0.148
No	134 (53.2%)	738 (58.1%)	
Yes	118 (46.8%)	532 (41.9%)	
Clergy appointment, n (%)			0.074
Pastoral charge	210 (82.4%)	1107 (86.6%)	
Extension or other	45 (17.6%)	171 (13.4%)	
Bi-vocational, n (%)			<0.001
No	246 (96.5%)	1115 (87.6%)	
Yes	9 (3.5%)	158 (12.4%)	
Hours per week worked as full-time clergy			0.521
Mean (SD)	49.5 (10.5)	49.9 (8.6)	
Financial stress, n (%)			0.089
Not at all or slightly stressful	172 (68.8%)	792 (63.2%)	
Moderately, very, or extremely	78 (31.2%)	462 (36.8%)	
Body Mass Index (BMI)			0.088
Mean (SD)	30.8 (7.0)	30.0 (6.7)	
Obesity, n (%)			0.043
Not obese (BMI <30)	131 (51.4%)	728 (58.2%)	
Obese (BMI 30+)	124 (48.6%)	522 (41.8%)	
Hypertension, n (%)			0.076
No (including missing)	167 (65.5%)	761 (59.5%)	

Yes, current or history	88 (34.5%)	517 (40.5%)	
Diabetes, n (%)			0.002
No (including missing)	223 (87.5%)	1012 (79.2%)	
Yes, current or history	32 (12.5%)	266 (20.8%)	
PHQ-8 depression symptoms sum score, [0-24]			<0.001
Mean (SD)	5.4 (4.6)	4.2 (4.1)	
Depression screens, n (%)			0.005
Negative (PHQ8 <10)	210 (83.0%)	1133 (89.3%)	
Positive (PHQ8 10+)	43 (17.0%)	136 (10.7%)	

[1]Comparison data are from the Clergy Health Longitudinal Survey 2019 wave (73% response rate), of the same study population from which the Selah Study recruited. Comparison participants were appointed and actively serving as clergy at the time of the survey.

[2]*P*-values generated using Kruskal Wallis test for continuous variables chi square text for categorical variables.

Supplemental Table B2. Baseline characteristics of immediate intervention and waitlist study arms for trial phase participants, for qualified individuals consenting to give HRV data

	Waitlist	Stress-Proofing	Daily Examen	Mindfulness-based Stress Reduction	Total
	(N = 40)	(N = 32)	(N = 38)	(N = 41)	(N = 151)
Age (in years)					
Mean (SD)	53.6 (10.5)	53.6 (10.5)	53.8 (12.4)	52.9 (11.4)	53.5 (11.2)
Sex, n (%)					
Female	18 (45.0%)	21 (65.6%)	17 (44.7%)	20 (48.8%)	76 (50.3%)
Male	22 (55.0%)	11 (34.4%)	21 (55.3%)	21 (51.2%)	75 (49.7%)
Race/Ethnicity, n (%)					
White and not Latinx	36 (90.0%)	30 (93.8%)	36 (94.7%)	39 (95.1%)	141 (93.4%)
African American and not Latinx	3 (7.5%)	2 (6.3%)	1 (2.6%)	0 (0.0%)	6 (4.0%)
Asian American/Pacific Islander, Native American, Latinx, bi/multi-racial, and other	1 (2.5%)	0 (0.0%)	1 (2.6%)	2 (4.9%)	4 (2.6%)
Marital & habitation status, n (%)					
Not married, separated or divorced	6 (15.0%)	6 (18.8%)	2 (5.3%)	2 (4.9%)	16 (10.6%)
Married or cohabitating	34 (85.0%)	26 (81.3%)	36 (94.7%)	39 (95.1%)	135 (89.4%)
Any children living at home, n (%)					
No	20 (50.0%)	19 (61.3%)	18 (47.4%)	15 (36.6%)	72 (48.0%)
Yes	20 (50.0%)	12 (38.7%)	20 (52.6%)	26 (63.4%)	78 (52.0%)
Clergy appointment, n (%)					
Pastoral charge	33 (82.5%)	24 (75.0%)	33 (86.8%)	35 (85.4%)	125 (82.8%)
Extension or other	7 (17.5%)	8 (25.0%)	5 (13.2%)	6 (14.6%)	26 (17.2%)
Bi-vocational, n (%)					
No	38 (95.0%)	31 (96.9%)	38 (100.0%)	39 (95.1%)	146 (96.7%)
Yes	2 (5.0%)	1 (3.1%)	0 (0.0%)	2 (4.9%)	5 (3.3%)
Hours per week worked as full-time clergy					
Mean (SD)	51.5 (11.8)	51.2 (10.6)	47.8 (6.3)	47.7 (11.0)	49.5 (10.3)
Stress from congregation(s)/work from Nov 2019 to registration, [0-3]					
Mean (SD)	1.8 (0.8)	2.2 (0.8)	1.8 (0.7)	1.8 (0.7)	1.9 (0.8)
Financial stress, n (%)					
Not at all or slightly stressful	31 (77.5%)	20 (64.5%)	29 (76.3%)	25 (62.5%)	105 (70.5%)
Moderately, very, or extremely	9 (22.5%)	11 (35.5%)	9 (23.7%)	15 (37.5%)	44 (29.5%)
Alcoholic drink intake, n (%)					
None	12 (30.0%)	12 (38.7%)	9 (24.3%)	12 (30.8%)	45 (30.6%)
Occasional drink (not every week)	12 (30.0%)	6 (19.4%)	11 (29.7%)	11 (28.2%)	40 (27.2%)
1-2 drinks	7 (17.5%)	6 (19.4%)	8 (21.6%)	8 (20.5%)	29 (19.7%)
3-6 drinks	6 (15.0%)	5 (16.1%)	4 (10.8%)	5 (12.8%)	20 (13.6%)
about a drink a day	2 (5.0%)	2 (6.5%)	3 (8.1%)	3 (7.7%)	10 (6.8%)
more than a drink a day	1 (2.5%)	0 (0.0%)	2 (5.4%)	0 (0.0%)	3 (2.0%)

Self-reported current heavy alcohol use, n (%)					
No	39 (100.0%)	30 (96.8%)	35 (94.6%)	38 (100.0%)	142 (97.9%)
Yes	0 (0.0%)	1 (3.2%)	2 (5.4%)	0 (0.0%)	3 (2.1%)
Caffeinated beverage intake per day, n (%)					
None	4 (10.0%)	5 (16.1%)	3 (8.1%)	5 (12.8%)	17 (11.6%)
1 cup	11 (27.5%)	5 (16.1%)	9 (24.3%)	10 (25.6%)	35 (23.8%)
2-3 cups	19 (47.5%)	14 (45.2%)	18 (48.6%)	19 (48.7%)	70 (47.6%)
4-5 cups	4 (10.0%)	6 (19.4%)	5 (13.5%)	5 (12.8%)	20 (13.6%)
6 or more cups	2 (5.0%)	1 (3.2%)	2 (5.4%)	0 (0.0%)	5 (3.4%)
Metabolic equivalents (METs) per week					
Mean (SD)	63.0 (68.8)	78.6 (74.9)	71.8 (89.5)	46.7 (71.8)	64.2 (76.7)
Body Mass Index (BMI)					
Mean (SD)	30.4 (6.6)	31.2 (7.7)	30.0 (6.4)	30.8 (7.9)	30.6 (7.1)
Obesity, n (%)					
Not obese (BMI <30)	21 (52.5%)	17 (53.1%)	20 (52.6%)	21 (51.2%)	79 (52.3%)
Obese (BMI 30+)	19 (47.5%)	15 (46.9%)	18 (47.4%)	20 (48.8%)	72 (47.7%)
High blood pressure, n (%)					
No (including missing)	24 (60.0%)	23 (71.9%)	23 (60.5%)	31 (75.6%)	101 (66.9%)
Yes, current or history	16 (40.0%)	9 (28.1%)	15 (39.5%)	10 (24.4%)	50 (33.1%)
Diabetes, n (%)					
No (including missing)	33 (82.5%)	29 (90.6%)	33 (86.8%)	38 (92.7%)	133 (88.1%)
Yes, current or history	7 (17.5%)	3 (9.4%)	5 (13.2%)	3 (7.3%)	18 (11.9%)
PHQ-8 depression symptoms sum score, [0-24]					
Mean (SD)	4.3 (4.3)	5.5 (4.3)	5.5 (4.5)	5.7 (4.8)	5.2 (4.5)
Depression screens, n (%)					
Negative (PHQ-8 <10)	36 (90.0%)	27 (84.4%)	32 (84.2%)	35 (85.4%)	130 (86.1%)
Positive (PHQ-8 10+)	4 (10.0%)	5 (15.6%)	6 (15.8%)	6 (14.6%)	21 (13.9%)

Supplemental Table B3. Baseline characteristics of pooled trial and observational participants

	Waitlist	Stress-Proofing	Daily Examen	Mindfulness-based Stress Reduction	Total
	(N = 73)	(N = 77)	(N = 97)	(N = 107)	(N = 354)
Age (in years)					
Mean (SD)	54.5 (10.3)	52.2 (11.7)	55.4 (11.4)	52.0 (11.9)	53.5 (11.4)
Sex, n (%)					
Female	32 (43.8%)	48 (62.3%)	41 (42.3%)	56 (52.3%)	177 (50.0%)
Male	41 (56.2%)	29 (37.7%)	56 (57.7%)	51 (47.7%)	177 (50.0%)
Race/Ethnicity, n (%)					
White and not Latinx	67 (91.8%)	64 (83.1%)	84 (86.6%)	100 (93.5%)	315 (89.0%)
African American and not Latinx	5 (6.8%)	7 (9.1%)	6 (6.2%)	4 (3.7%)	22 (6.2%)
Asian-American/Pacific Islander, Native American, Latinx, multi-racial, and other	1 (1.4%)	6 (7.8%)	7 (7.2%)	3 (2.8%)	17 (4.8%)
Marital & habitation status, n (%)					
Not married, or married but separated/divorcing	8 (11.0%)	16 (20.8%)	9 (9.3%)	12 (11.2%)	45 (12.7%)
Married or cohabitating	65 (89.0%)	61 (79.2%)	88 (90.7%)	95 (88.8%)	309 (87.3%)
Any children living at home, n (%)					
No	39 (53.4%)	43 (58.1%)	59 (61.5%)	53 (50.5%)	194 (55.7%)
Yes	34 (46.6%)	31 (41.9%)	37 (38.5%)	52 (49.5%)	154 (44.3%)
Clergy appointment, n (%)					
Pastoral charge	59 (80.8%)	61 (79.2%)	82 (84.5%)	85 (79.4%)	287 (81.1%)
Extension or other	14 (19.2%)	16 (20.8%)	15 (15.5%)	22 (20.6%)	67 (18.9%)
Bi-vocational, n (%)					
No	70 (95.9%)	71 (93.4%)	93 (95.9%)	104 (97.2%)	338 (95.8%)
Yes	3 (4.1%)	5 (6.6%)	4 (4.1%)	3 (2.8%)	15 (4.2%)
Hours per week worked as full-time clergy					
Mean (SD)	49.4 (9.9)	48.7 (10.8)	49.5 (10.6)	49.0 (10.2)	49.2 (10.3)
Stress from congregation(s)/work from Nov 2019 to registration, [0-3]					
Mean (SD)	1.8 (0.7)	2.0 (0.8)	1.8 (0.6)	1.9 (0.7)	1.9 (0.7)
Financial stress, n (%)					
Not at all or slightly stressful	50 (69.4%)	48 (64.9%)	68 (70.1%)	71 (68.3%)	237 (68.3%)
Moderately, very, or extremely	22 (30.6%)	26 (35.1%)	29 (29.9%)	33 (31.7%)	110 (31.7%)
Alcoholic drink intake, n (%)					
None	23 (31.9%)	21 (28.4%)	36 (37.5%)	33 (32.0%)	113 (32.8%)
Occasional drink (not every week)	17 (23.6%)	25 (33.8%)	23 (24.0%)	28 (27.2%)	93 (27.0%)
1-2 drinks	11 (15.3%)	11 (14.9%)	17 (17.7%)	17 (16.5%)	56 (16.2%)
3-6 drinks	13 (18.1%)	11 (14.9%)	8 (8.3%)	16 (15.5%)	48 (13.9%)
about a drink a day	6 (8.3%)	4 (5.4%)	8 (8.3%)	8 (7.8%)	26 (7.5%)
more than a drink a day	2 (2.8%)	2 (2.7%)	4 (4.2%)	1 (1.0%)	9 (2.6%)

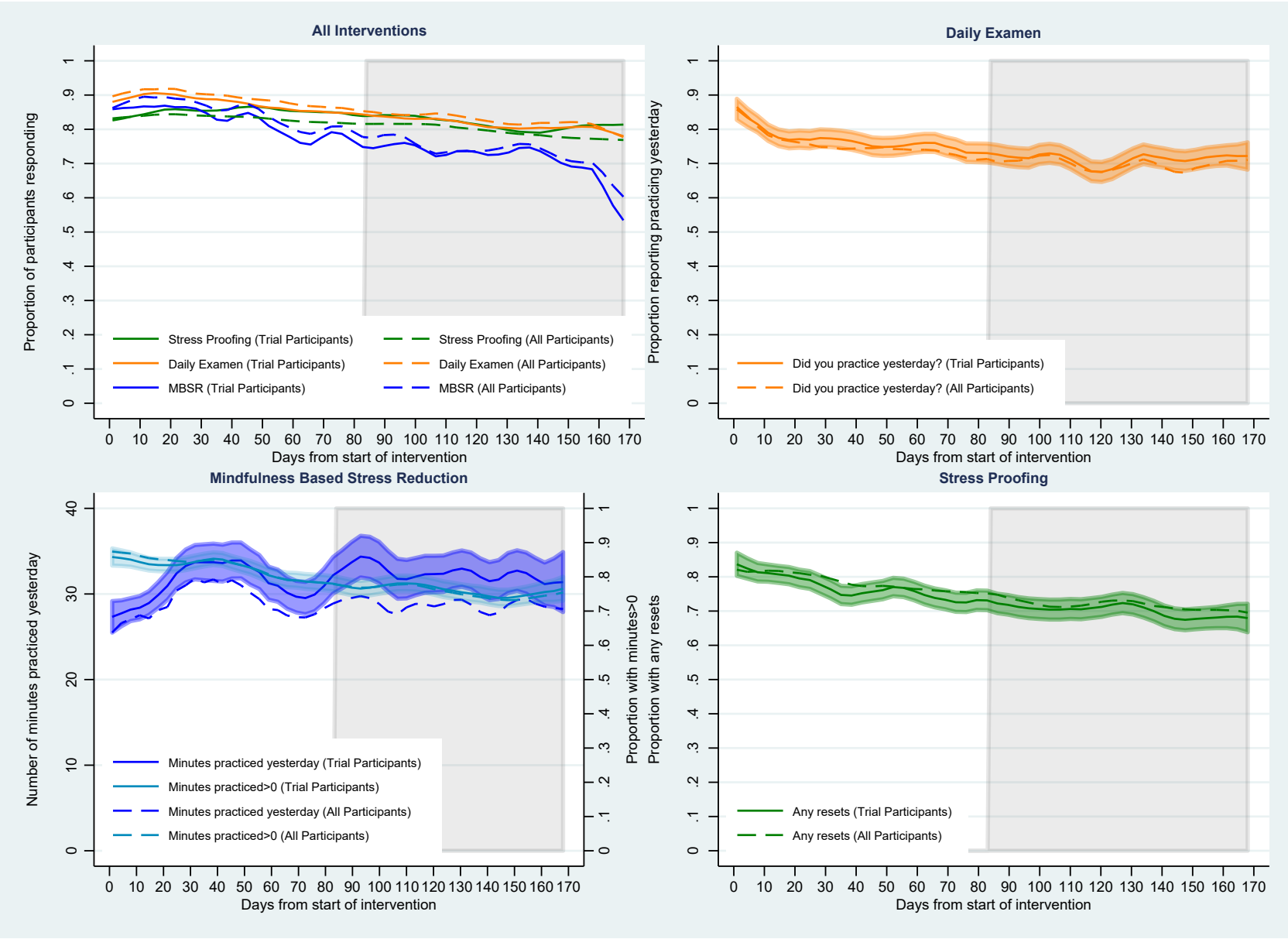
Self-reported current heavy alcohol use, n (%)					
No	70 (100.0%)	71 (97.3%)	93 (96.9%)	101 (99.0%)	335 (98.2%)
Yes	0 (0.0%)	2 (2.7%)	3 (3.1%)	1 (1.0%)	6 (1.8%)
Caffeinated beverage intake per day, n (%)					
None	9 (12.5%)	7 (9.5%)	12 (12.5%)	11 (10.7%)	39 (11.3%)
1 cup	18 (25.0%)	21 (28.4%)	16 (16.7%)	30 (29.1%)	85 (24.6%)
2-3 cups	33 (45.8%)	33 (44.6%)	54 (56.3%)	46 (44.7%)	166 (48.1%)
4-5 cups	10 (13.9%)	12 (16.2%)	10 (10.4%)	12 (11.7%)	44 (12.8%)
6 or more cups	2 (2.8%)	1 (1.4%)	4 (4.2%)	4 (3.9%)	11 (3.2%)
Metabolic equivalents (METs) per week					
Mean (SD)	70.0 (89.1)	53.6 (64.6)	80.1 (99.7)	50.7 (68.1)	63.4 (82.3)
Body Mass Index (BMI)					
Mean (SD)	30.4 (6.9)	31.0 (6.7)	30.4 (6.5)	30.9 (7.4)	30.7 (6.9)
Obesity, n (%)					
Not obese (BMI <30)	39 (53.4%)	34 (44.2%)	49 (50.5%)	55 (51.4%)	177 (50.0%)
Obese (BMI 30+)	34 (46.6%)	43 (55.8%)	48 (49.5%)	52 (48.6%)	177 (50.0%)
High blood pressure, n (%)					
No (including missing)	45 (61.6%)	52 (67.5%)	63 (64.9%)	77 (72.0%)	237 (66.9%)
Yes, current or history	28 (38.4%)	25 (32.5%)	34 (35.1%)	30 (28.0%)	117 (33.1%)
Diabetes, n (%)					
No (including missing)	60 (82.2%)	72 (93.5%)	81 (83.5%)	99 (92.5%)	312 (88.1%)
Yes, current or history	13 (17.8%)	5 (6.5%)	16 (16.5%)	8 (7.5%)	42 (11.9%)
PHQ-8 depression symptoms sum score, [0-24]					
Mean (SD)	4.2 (3.9)	5.8 (4.6)	5.1 (4.3)	6.4 (5.2)	5.4 (4.6)
Depression screens, n (%)					
Negative (PHQ-8 <10)	64 (88.9%)	59 (76.6%)	83 (85.6%)	80 (75.5%)	286 (81.3%)
Positive (PHQ-8 10+)	8 (11.1%)	18 (23.4%)	14 (14.4%)	26 (24.5%)	66 (18.8%)

Note. Of the sample of 354, some participants had dual baselines; there were 307 total unique participants (73 were initially assigned to the waitlist, 68 to Stress Proofing, 82 to Daily Examen, and 84 to Mindfulness-based Stress Reduction)

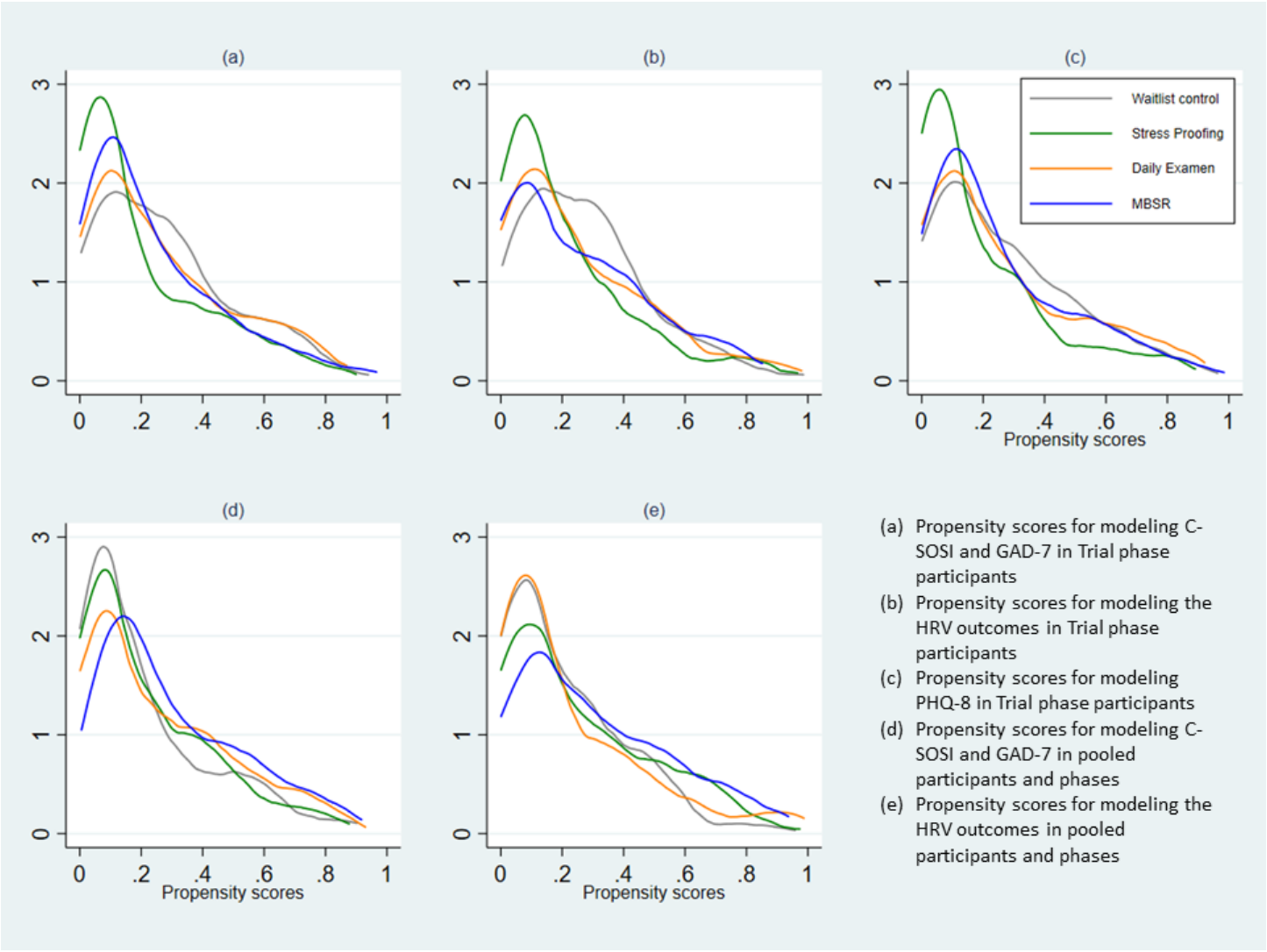
Supplemental Table B4. Intervention attendance and participation

	Trial participants (N = 184)	Pooled Trial and Observational Participants (N = 234)
Number of Stress Proofing main sessions attended (Range 0-4)		
0	0 (0.0%)	2 (2.9%)
1	2 (4.2%)	3 (4.4%)
2	4 (8.3%)	6 (8.8%)
3	12 (25.0%)	14 (20.6%)
4	30 (62.5%)	43 (63.2%)
Did the participant attend the Stress Proofing follow-up session?		
No	18 (42.9%)	27 (46.6%)
Yes	24 (57.1%)	31 (53.4%)
Number of Daily Examen main sessions attended (Range 0-3)		
0	0 (0.0%)	0 (0.0%)
1	0 (0.0%)	0 (0.0%)
2	3 (4.2%)	4 (4.9%)
3	68 (95.8%)	78 (95.1%)
Number of Daily Examen follow-up sessions attended (Range 0-2)		
0	34 (47.9%)	39 (47.6%)
1	18 (25.4%)	23 (28.0%)
2	19 (26.8%)	20 (24.4%)
Number of Mindfulness-based Stress Reduction main sessions attended (Range 0-8)		
Mean (SD)	6.9 (1.5)	7.0 (1.4)
Median (Q1, Q3)	7.0 (7.0, 8.0)	7.0 (7.0, 8.0)

Supplemental Figure B3. Text message response rates, practice responses, and trends over 24 weeks



Supplemental Figure B4. Distribution of propensity scores by treatment condition, outcome, and population type



Supplemental Table B5. Baseline characteristic balance between treatment conditions before and after propensity score adjustment for C-SOSI and anxiety outcomes in trial participants

	Unadjusted Mean (SD) or % (n)				*P-value*	
	Stress Proofing	Daily Examen	Mindfulness-based Stress Reduction	Waitlist	Before adjustment	After adjustment
Age (in years), mean (SD)	53.2 (10.7)	54.0 (12.1)	52.7 (11.7)	54.7 (10.1)	.784	.854
Sex, n (%)					.169	.335
Male	37.0% (17)	58.7% (37)	51.7% (30)	51.5% (34)		
Female	63.0% (29)	41.3% (26)	48.3% (28)	48.5% (32)		
Race & ethnicity, n (%)					.604	.826
Single-racial, non-Latinx white	93.5% (43)	87.3% (55)	93.1% (54)	92.4% (61)		
All other ethnicities, including bi-/multi-racial	6.5% (3)	12.7% (8)	6.9% (4)	7.6% (5)		
Marital & habitation status, n (%)					.157	.366
Not married, or married but separated/divorcing	17.4% (8)	6.3% (4)	5.2% (3)	12.1% (8)		
Married or cohabitating with partner	82.6% (38)	93.7% (59)	94.8% (55)	87.9% (58)		
Any children living at home, n (%)					.374	.705
No	58.7% (27)	54.0% (34)	43.1% (25)	56.1% (37)		
Yes	41.3% (19)	46.0% (29)	56.9% (33)	43.9% (29)		
Full-time clergy, n (%)					.460	.745
No	15.2% (7)	20.6% (13)	10.3% (6)	13.6% (9)		
Yes	84.8% (39)	79.4% (50)	89.7% (52)	86.4% (57)		
Bi-vocational, n (%)					.816	.991
No	95.7% (44)	98.4% (62)	96.6% (56)	95.5% (63)		
Yes	4.3% (2)	1.6% (1)	3.4% (2)	4.5% (3)		
Number of congregations appointed to, n (%)					.589	.704
Not appointed to a local congregation	17.4% (8)	12.7% (8)	15.5% (9)	16.7% (11)		
1 congregation	52.2% (24)	65.1% (41)	63.8% (37)	68.2% (45)		
2+ congregations	30.4% (14)	22.2% (14)	20.7% (12)	15.2% (10)		
Number of congregants pastored, n (%)					.501	.795
Not appointed to a local congregation	19.6% (9)	15.9% (10)	20.7% (12)	18.2% (12)		

1-149 people in worship per week	60.9% (28)	68.3% (43)	60.3% (35)	51.5% (34)		
150+ people in worship per week	19.6% (9)	15.9% (10)	19.0% (11)	30.3% (20)		
Hours per week worked as clergy, mean (SD)	46.5 (13.8)	44.7 (13.6)	47.1 (13.5)	47.2 (11.8)	.686	.747
Alcoholic drink intake, scale range 0-5, mean (SD)	1.35 (1.35)	1.24 (1.34)	1.47 (1.39)	1.50 (1.46)	.711	.900
Caffeinated beverage intake, scale range 0-4, mean (SD)	1.78 (0.89)	1.67 (0.95)	1.76 (1.00)	1.70 (0.96)	.913	.990
Metabolic equivalents (METs) per week, mean (SD)	58.4 (62.6)	74.6 (85.7)	58.4 (77.8)	70.8 (91.8)	.609	.979
Practicing Daily Examen 3+ times a week at registration, n (%)					.883	.937
No	93.5% (43)	90.5% (57)	93.1% (54)	93.9% (62)		
Yes	6.5% (3)	9.5% (6)	6.9% (4)	6.1% (4)		
Practicing Mindfulness 3+ times a week at registration, n (%)					.532	.711
No or missing	93.5% (43)	90.5% (57)	96.6% (56)	95.5% (63)		
Yes	6.5% (3)	9.5% (6)	3.4% (2)	4.5% (3)		
Number of top preferences among Selah interventions, n (%)					.296	.817
No preference	21.7% (10)	25.4% (16)	20.7% (12)	12.1% (8)		
1 top preference	76.1% (35)	68.3% (43)	69.0% (40)	81.8% (54)		
2 tied preferences	2.2% (1)	6.3% (4)	10.3% (6)	6.1% (4)		
Body Mass Index, mean (SD)	31.0 (7.0)	29.9 (6.4)	32.0 (8.0)	30.4 (7.1)	.431	.987
Self-endorsed high cholesterol, n (%)					.188	.823
Never or missing	73.9% (34)	54.0% (34)	56.9% (33)	59.1% (39)		
Yes, current or history	26.1% (12)	46.0% (29)	43.1% (25)	40.9% (27)		
Overall life stress at registration, scale range 0-4, mean (SD)	2.46 (0.75)	2.32 (0.78)	2.45 (0.82)	2.42 (0.91)	.781	.972
C-SOSI stress symptoms, scale range 0-4, mean (SD)	1.12 (0.60)	0.93 (0.56)	1.16 (0.54)	0.87 (0.52)	.010	.498
Self-endorsed single-item anxiety, n (%)					.016	.538
Never or missing	63.0% (29)	77.8% (49)	51.7% (30)	72.7% (48)		
Yes, current or history	37.0% (17)	22.2% (14)	48.3% (28)	27.3% (18)		
GAD-7 anxiety symptoms, scale range 0-21, mean (SD)	5.0 (4.0)	4.6 (4.9)	6.0 (5.1)	4.2 (3.9)	.133	.713
Self-endorsed single-item depression, n (%)					.015	.580
Never or missing	60.9% (28)	73.0% (46)	46.6% (27)	69.7% (46)		
Yes, current or history	39.1% (18)	27.0% (17)	53.4% (31)	30.3% (20)		

Supplemental Table B6. Baseline characteristic balance between treatment conditions before and after propensity score adjustment for HRV outcomes in trial participants

	Unadjusted Mean (SD) or % (n)				*P-value*	
	Stress Proofing	Daily Examen	Mindfulness-Based Stress Reduction	Waitlist	Before adjustment	After adjustment
Age (in years), mean (SD)	53.2 (10.7)	53.5 (12.6)	53.3 (10.8)	53.4 (10.5)	1.000	.950
Sex, n (%)					.281	.781
Male	33.3% (10)	55.6% (20)	48.6% (18)	53.8% (21)		
Female	66.7% (20)	44.4% (16)	51.4% (19)	46.2% (18)		
Race & ethnicity, n (%)					.832	.994
Single-racial, non-Hispanic White	93.3% (28)	94.4% (34)	94.6% (35)	89.7% (35)		
All other ethnicities, including bi-/multi-racial	6.7% (2)	5.6% (2)	5.4% (2)	10.3% (4)		
Marital & habitation status, n (%)					.118	.960
Not married, or married but separated/divorcing	20.0% (6)	5.6% (2)	2.7% (1)	15.4% (6)		
Married or cohabitating with partner	80.0% (24)	94.4% (34)	97.3% (36)	84.6% (33)		
Any children at home, n (%)					.101	.856
No	60.0% (18)	47.2% (17)	29.7% (11)	48.7% (19)		
Yes	40.0% (12)	52.8% (19)	70.3% (26)	51.3% (20)		
Full-time clergy in UMC, n (%)					.810	.987
No	16.7% (5)	22.2% (8)	13.5% (5)	17.9% (7)		
Yes	83.3% (25)	77.8% (28)	86.5% (32)	82.1% (32)		
Number of congregations appointed to, n (%)					.419	.968
Not appointed to a local congregation	23.3% (7)	11.1% (4)	10.8% (4)	15.4% (6)		
1 congregation	43.3% (13)	69.4% (25)	67.6% (25)	64.1% (25)		
2+ congregations	33.3% (10)	19.4% (7)	21.6% (8)	20.5% (8)		
Number of congregants pastored, n (%)					.220	.970
Not appointed to a local congregation	26.7% (8)	13.9% (5)	16.2% (6)	17.9% (7)		
1-149 people in worship per week	60.0% (18)	69.4% (25)	64.9% (24)	46.2% (18)		
150+ people in worship per week	13.3% (4)	16.7% (6)	18.9% (7)	35.9% (14)		
Hours per week worked as UMC clergy, mean (SD)	47.6 (14.4)	42.6 (11.7)	45.3 (13.2)	47.7 (14.3)	.328	.967

Alcoholic drink intake, scale range 0-5, mean (SD)	1.30 (1.34)	1.61 (1.46)	1.35 (1.27)	1.38 (1.33)	.783	.850
Caffeinated beverage intake, scale range 0-4, mean (SD)	1.83 (1.02)	1.83 (0.97)	1.59 (0.90)	1.72 (0.97)	.687	.957
Metabolic equivalents (METs) per week, mean (SD)	74.3 (65.2)	75.8 (90.3)	47.4 (73.1)	61.2 (68.7)	.352	.952
Practicing Daily Examen 3+ times a week at registration, n (%)						
No	96.7% (29)	94.4% (34)	97.3% (36)	94.9% (37)		
Yes	3.3% (1)	5.6% (2)	2.7% (1)	5.1% (2)		
Practicing Mindfulness 3+ times a week at registration, n (%)					.333	.942
No or missing	96.7% (29)	88.9% (32)	97.3% (36)	97.4% (38)		
Yes	3.3% (1)	11.1% (4)	2.7% (1)	2.6% (1)		
Number of top preferences among Selah interventions, n (%)					.768	1.000
No preference	23.3% (7)	19.4% (7)	27.0% (10)	17.9% (7)		
1 top preference	73.3% (22)	72.2% (26)	64.9% (24)	79.5% (31)		
2 tied preferences	3.3% (1)	8.3% (3)	8.1% (3)	2.6% (1)	.768	1.000
Body Mass Index, mean (SD)	31.2 (8.0)	29.6 (6.1)	31.3 (8.0)	30.5 (6.6)	.714	.985
Self-endorsed high cholesterol, n (%)						
Never or missing	76.7% (23)	58.3% (21)	59.5% (22)	56.4% (22)		
Yes, current or history	23.3% (7)	41.7% (15)	40.5% (15)	43.6% (17)		
Overall life stress at registration, scale range 0-4, mean (SD)	2.50 (0.78)	2.39 (0.84)	2.35 (0.86)	2.44 (1.05)	.916	.983
C-SOSI stress symptoms, scale range 0-4, mean (SD)	1.14 (0.63)	0.98 (0.55)	1.07 (0.44)	0.87 (0.57)	.199	.863
Self-endorsed anxiety, n (%)					.118	.893
Never or missing	60.0% (18)	72.2% (26)	48.6% (18)	71.8% (28)		
Yes, current or history	40.0% (12)	27.8% (10)	51.4% (19)	28.2% (11)		
GAD7 anxiety symptoms, scale range 0-21, mean (SD)	5.0 (3.8)	4.7 (4.7)	5.2 (4.7)	4.2 (4.0)	.782	.981
Self-endorsed depression, n (%)					.345	.998
Never or missing	60.0% (18)	66.7% (24)	48.6% (18)	66.7% (26)		
Yes, current or history	40.0% (12)	33.3% (12)	51.4% (19)	33.3% (13)		

Supplemental Table B7. Baseline characteristic balance between treatment conditions before and after p-value adjustment for C-SOSI and anxiety outcomes in pooled trial and observational participants

	Unadjusted Mean (SD) or % (n)				*P-value*	
	Stress Proofing	Daily Examen	Mindfulness Based Stress Reduction	Waitlist	Before adjustment	After adjustment
Age (in years), mean (SD)	51.2 (11.3)	55.6 (11.9)	52.5 (11.1)	54.7 (10.1)	.030	.417
Sex, n (%)					.055	.245
Male	37.2% (29)	57.0% (57)	43.8% (49)	49.1% (52)		
Female	62.8% (49)	43.0% (43)	56.3% (63)	50.9% (54)		
Race & ethnicity, n (%)					.227	.421
Single-racial, non-Latinx white	87.2% (68)	88.0% (88)	94.6% (106)	92.5% (98)		
All other ethnicities, including bi-/multi-racial	12.8% (10)	12.0% (12)	5.4% (6)	7.5% (8)		
Marital & habitation status, n (%)					.318	.485
Not married, or married but separated/divorcing	17.9% (14)	9.0% (9)	10.7% (12)	12.3% (13)		
Married or cohabitating with partner	82.1% (64)	91.0% (91)	89.3% (100)	87.7% (93)		
Any children living at home, n (%)					.131	.473
No	59.0% (46)	63.0% (63)	47.3% (53)	55.7% (59)		
Yes	41.0% (32)	37.0% (37)	52.7% (59)	44.3% (47)		
Full-time clergy, n (%)					.201	.390
No	15.4% (12)	21.0% (21)	10.7% (12)	13.2% (14)		
Yes	84.6% (66)	79.0% (79)	89.3% (100)	86.8% (92)		
Bi-vocational, n (%)					.800	.792
No	94.9% (74)	95.0% (95)	97.3% (109)	96.2% (102)		
Yes	5.1% (4)	5.0% (5)	2.7% (3)	3.8% (4)		
Number of congregations appointed to, n (%)					.556	.644
Not appointed to a local congregation	16.7% (13)	12.0% (12)	20.5% (23)	16.0% (17)		
1 congregation	65.4% (51)	69.0% (69)	62.5% (70)	71.7% (76)		
2+ congregations	17.9% (14)	19.0% (19)	17.0% (19)	12.3% (13)		
Number of congregants pastored, n (%)					.243	.403
Not appointed to a local congregation	20.5% (16)	15.0% (15)	23.2% (26)	17.9% (19)		

1-149 people in worship per week	55.1% (43)	67.0% (67)	56.3% (63)	51.9% (55)		
150+ people in worship per week	24.4% (19)	18.0% (18)	20.5% (23)	30.2% (32)		
Hours per week worked as clergy, mean (SD)	45.0 (14.1)	44.4 (13.3)	46.9 (12.8)	47.2 (11.8)	.334	.475
Alcoholic drink intake, scale range 0-5, mean (SD)	1.49 (1.37)	1.21 (1.28)	1.45 (1.37)	1.54 (1.47)	.344	.509
Caffeinated beverage intake, scale range 0-4, mean (SD)	1.71 (0.88)	1.78 (0.96)	1.65 (0.97)	1.70 (0.97)	.806	.762
Metabolic equivalents (METs) per week, mean (SD)	55.6 (59.7)	85.7 (105.5)	46.1 (65.0)	73.3 (97.2)	.004	.213
Practicing Daily Examen 3+ times a week at registration, n (%)					.579	.648
No	91.0% (71)	92.0% (92)	95.5% (107)	94.3% (100)		
Yes	9.0% (7)	8.0% (8)	4.5% (5)	5.7% (6)		
Practicing Mindfulness 3+ times a week at registration, n (%)					.468	.603
No or missing	92.3% (72)	92.0% (92)	96.4% (108)	95.3% (101)		
Yes	7.7% (6)	8.0% (8)	3.6% (4)	4.7% (5)		
Number of top preferences among Selah interventions, n (%)					.429	.985
No preference	21.8% (17)	18.0% (18)	15.2% (17)	11.3% (12)		
1 top preference	74.4% (58)	74.0% (74)	75.9% (85)	82.1% (87)		
2 tied preferences	3.8% (3)	8.0% (8)	8.9% (10)	6.6% (7)		
Body Mass Index, mean (SD)	31.2 (7.3)	30.0 (6.6)	31.1 (7.3)	30.3 (7.2)	.615	.975
Self-endorsed high cholesterol, n (%)					.036	.405
Never or missing	73.1% (57)	52.0% (52)	60.7% (68)	56.6% (60)		
Yes, current or history	26.9% (21)	48.0% (48)	39.3% (44)	43.4% (46)		
Overall life stress at registration, scale range 0-4, mean (SD)	2.49 (0.75)	2.31 (0.84)	2.46 (0.88)	2.43 (0.88)	.474	.897
C-SOSI stress symptoms, scale range 0-4, mean (SD)	1.15 (0.57)	0.88 (0.55)	1.03 (0.59)	0.88 (0.52)	.003	.156
Self-endorsed single-item anxiety, n (%)					.016	.374
Never or missing	62.8% (49)	78.0% (78)	59.8% (67)	73.6% (78)		
Yes, current or history	37.2% (29)	22.0% (22)	40.2% (45)	26.4% (28)		
GAD-7 anxiety symptoms, scale range 0-21, mean (SD)	5.2 (4.0)	4.1 (4.5)	5.1 (4.8)	4.2 (3.9)	.173	.651
Self-endorsed single item depression, n (%)					.001	.101
Never or missing	60.3% (47)	76.0% (76)	50.9% (57)	68.9% (73)		
Yes, current or history	39.7% (31)	24.0% (24)	49.1% (55)	31.1% (33)		

Supplemental Table B8. Baseline characteristic balance between treatment conditions before and after propensity score adjustment for HRV outcomes in pooled trial and observational participants

	Unadjusted Mean (SD) or % (n)				*P-value*	
	Stress Proofing	Daily Examen	Mindfulness-Based Stress Reduction	Waitlist	Before adjustment	After adjustment
Age (in years), mean (SD)	51.2 (10.9)	53.5 (12.3)	52.9 (10.4)	52.9 (10.1)	.784	.920
Sex, n (%)					.108	.256
Male	31.1% (14)	55.6% (25)	44.6% (29)	51.0% (26)		
Female	68.9% (31)	44.4% (20)	55.4% (36)	49.0% (25)		
Race & ethnicity, n (%)					.878	.895
Single-racial, non-Latinx white	93.3% (42)	91.1% (41)	93.8% (61)	90.2% (46)		
All other ethnicities, including bi-/multi-racial	6.7% (3)	8.9% (4)	6.2% (4)	9.8% (5)		
Marital & habitation status, n (%)					.741	.786
Not married, or married but separated/divorcing	15.6% (7)	8.9% (4)	12.3% (8)	15.7% (8)		
Married or cohabitating with partner	84.4% (38)	91.1% (41)	87.7% (57)	84.3% (43)		
Any children living at home, n (%)					.722	.914
No	53.3% (24)	51.1% (23)	43.1% (28)	47.1% (24)		
Yes	46.7% (21)	48.9% (22)	56.9% (37)	52.9% (27)		
Full-time clergy in UMC, n (%)					.761	.990
No	15.6% (7)	17.8% (8)	10.8% (7)	13.7% (7)		
Yes	84.4% (38)	82.2% (37)	89.2% (58)	86.3% (44)		
Number of congregations appointed to, n (%)					.577	.571
Not appointed to a local congregation	24.4% (11)	11.1% (5)	15.4% (10)	17.6% (9)		
1 congregation	53.3% (24)	73.3% (33)	67.7% (44)	66.7% (34)		
2+ congregations	22.2% (10)	15.6% (7)	16.9% (11)	15.7% (8)		
Number of congregants pastored, n (%)					.116	.438
Not appointed to a local congregation	31.1% (14)	13.3% (6)	18.5% (12)	19.6% (10)		
1-149 people in worship per week	51.1% (23)	68.9% (31)	58.5% (38)	45.1% (23)		
150+ people in worship per week	17.8% (8)	17.8% (8)	23.1% (15)	35.3% (18)		
Hours per week worked as clergy, mean (SD)	47.3 (12.6)	44.6 (12.3)	46.7 (12.7)	49.2 (13.9)	.392	.923

Alcoholic drink intake, scale range 0-5, mean (SD)	1.42 (1.25)	1.40 (1.44)	1.32 (1.32)	1.31 (1.32)	.969	.956
Caffeinated beverage intake, scale range 0-4, mean (SD)	1.78 (0.88)	1.73 (0.96)	1.57 (0.87)	1.65 (0.91)	.635	.663
Metabolic equivalents (METs) per week, mean (SD)	75.1 (63.0)	72.7 (100.6)	39.1 (61.6)	64.6 (66.9)	.036	.670
Practicing Daily Examen 3+ times a week at registration, n (%)						
No	97.8% (44)	95.6% (43)	98.5% (64)	96.1% (49)		
Yes	2.2% (1)	4.4% (2)	1.5% (1)	3.9% (2)		
Practicing Mindfulness 3+ times a week at registration, n (%)					.102	.403
No or missing	97.8% (44)	86.7% (39)	96.9% (63)	96.1% (49)		
Yes	2.2% (1)	13.3% (6)	3.1% (2)	3.9% (2)		
Number of top preferences among Selah interventions, n (%)					.912	.990
No preference	20.0% (9)	20.0% (9)	20.0% (13)	17.6% (9)		
1 top preference	77.8% (35)	73.3% (33)	75.4% (49)	80.4% (41)		
2 tied preferences	2.2% (1)	6.7% (3)	4.6% (3)	2.0% (1)	.912	.990
Body Mass Index, mean (SD)	30.4 (7.1)	29.6 (6.3)	31.1 (7.8)	30.3 (6.7)	.774	.936
Self-endorsed high cholesterol, n (%)						
Never or missing	75.6% (34)	60.0% (27)	63.1% (41)	58.8% (30)		
Yes, current or history	24.4% (11)	40.0% (18)	36.9% (24)	41.2% (21)		
Overall life stress at registration, scale range 0-4, mean (SD)	2.56 (0.78)	2.58 (0.89)	2.45 (0.92)	2.57 (1.04)	.852	.953
C-SOSI stress symptoms, scale range 0-4, mean (SD)	1.11 (0.56)	0.98 (0.62)	0.96 (0.53)	0.86 (0.58)	.173	.477
Self-endorsed single-item anxiety, n (%)					.359	.846
Never or missing	64.4% (29)	73.3% (33)	58.5% (38)	70.6% (36)		
Yes, current or history	35.6% (16)	26.7% (12)	41.5% (27)	29.4% (15)		
GAD-7 anxiety symptoms, scale range 0-21, mean (SD)	5.0 (3.7)	4.7 (5.0)	4.5 (4.3)	4.2 (4.1)	.821	.894
Self-endorsed single-item depression, n (%)					.163	.982
Never or missing	62.2% (28)	68.9% (31)	49.2% (32)	64.7% (33)		
Yes, current or history	37.8% (17)	31.1% (14)	50.8% (33)	35.3% (18)		

Supplemental Table B9. Between arm, mixed effects regression estimated differences in outcomes between immediate intervention and waitlist control by follow-up time point for trial period participants, using estimates from multiply imputed data (Sensitivity Analysis)

	Time Point	Stress Proofing	Daily Examen	Mindfulness-based Stress Reduction
Survey outcomes				
C-SOSI (scale range 0-4)	12 weeks	-0.26 [-0.38, -0.13]	-0.08 [-0.22, 0.05]	-0.29 [-0.42, -0.16]
	24 weeks	-0.22 [-0.42, -0.02]	-0.20 [-0.36, -0.03]	-0.34 [-0.49, -0.20]
GAD-7 (scale range 0-21)	12 weeks	-1.27 [-2.30, -0.24]	-0.49 [-1.47, 0.50]	-1.71 [-2.65, -0.78]
	24 weeks	-1.11 [-2.71, 0.50]	-1.02 [-2.24, 0.20]	-1.76 [-2.85, -0.66]
PHQ-8 (scale range 0-24)	12 weeks	-1.83 [-3.33, -0.33]	-1.25 [-2.60, 0.10]	-1.95 [-3.15, -0.75]
	24 weeks	-0.98 [-2.97, 1.02]	-1.18 [-2.67, 0.32]	-2.06 [-3.51, -0.61]
HRV outcomes				
MESOR (unit = 1 millisecond)	12 weeks	0.52 [-3.39, 4.43]	1.62 [-2.44, 5.68]	3.80 [0.50, 7.10]
Amplitude (unit = 1 millisecond)	12 weeks	1.33 [-1.41, 4.06]	1.34 [-0.78, 3.47]	2.20 [-0.17, 4.56]

Abbrev: C-SOSI = Calgary Symptoms of Stress Inventory; GAD-7 = Generalized Anxiety Disorder-7; PHQ-8 = Patient Health Questionnaire-8; HRV = Heart rate variability; MESOR = Midline estimating statistic of rhythm

Supplemental Table B10. Between arm, mixed effects regression estimated differences in outcomes between immediate intervention and waitlist control by follow-up time point for trial period participants, using pooled trial period and observational data (Sensitivity Analysis)

	Sample size			Stress Proofing		Daily Examen		Mindfulness-Based Stress Reduction	
	Participants	Observations	Time Point	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Survey outcomes									
C-SOSI (scale range 0-4)	276	780	12 weeks	-0.24 [-0.34, -0.15]	-0.26 [-0.37, -0.16]	-0.06 [-0.19, 0.07]	-0.07 [-0.20, 0.06]	-0.25 [-0.35, -0.16]	-0.27 [-0.36, -0.18]
			24 weeks	-0.25 [-0.41, -0.10]	-0.35 [-0.49, -0.21]	-0.15 [-0.32, 0.02]	-0.21 [-0.36, -0.06]	-0.30 [-0.42, -0.17]	-0.36 [-0.49, -0.24]
GAD-7 (scale range 0-21)	276	780	12 weeks	-1.20 [-2.04, -0.36]	-1.44 [-2.34, -0.55]	-0.55 [-1.44, 0.35]	-0.60 [-1.49, 0.28]	-1.64 [-2.40, -0.88]	-1.76 [-2.53, -0.99]
			24 weeks	-1.13 [-2.44, 0.18]	-1.85 [-3.03, -0.66]	-0.87 [-2.07, 0.32]	-1.20 [-2.31, -0.09]	-1.78 [-2.81, -0.74]	-2.18 [-3.24, -1.11]
PHQ-8 (scale range 0-24)	275	777	12 weeks	-1.53 [-2.58, -0.47]	-1.71 [-2.84, -0.59]	-1.15 [-2.32, 0.02]	-1.14 [-2.30, 0.03]	-2.06 [-3.10, -1.01]	-2.18 [-3.23, -1.13]
			24 weeks	-1.65 [-2.74, -0.56]	-2.16 [-3.32, -1.00]	-0.98 [-2.23, 0.28]	-1.19 [-2.45, 0.07]	-1.94 [-2.93, -0.95]	-2.31 [-3.46, -1.17]
HRV outcomes									
MESOR (unit = 1 millisecond)	170	324	12 weeks	-1.91 [-4.99, 1.17]	-0.53 [-3.58, 2.53]	1.03 [-3.07, 5.14]	1.31 [-2.54, 5.15]	2.85 [-0.26, 5.95]	3.40 [0.43, 6.36]
Amplitude (unit = 1 millisecond)	170	324	12 weeks	0.76 [-1.31, 2.83]	1.05 [-1.09, 3.19]	1.48 [-0.40, 3.37]	1.48 [-0.27, 3.23]	2.14 [0.32, 3.96]	2.24 [0.32, 4.16]

Depending on the survey outcome, 70 Stress Proofing participants, 85 Daily Examen participants, and 94-95 MBSR participants were compared to 66 waitlist control participants. For each of the HRV outcomes, 42 Stress Proofing participants, 42 Daily Examen participants, and 59 MBSR participants were compared to 39 waitlist control participants. These participant numbers add up to be higher than the total number of participants in the main body of the table because participants who provided data both while on the waitlist and during an intervention in the post-waitlist period are counted twice.

Abbrev: C-SOSI = Calgary Symptoms of Stress Inventory; GAD-7 = Generalized Anxiety Disorder-7; PHQ-8 = Patient Health Questionnaire-8; HRV = Heart rate variability; MESOR = Midline estimating statistic of rhythm

Supplemental Table B11. Baseline characteristics by unique preferences vs. no unique preference for trial participants and pooled trial/observational participants

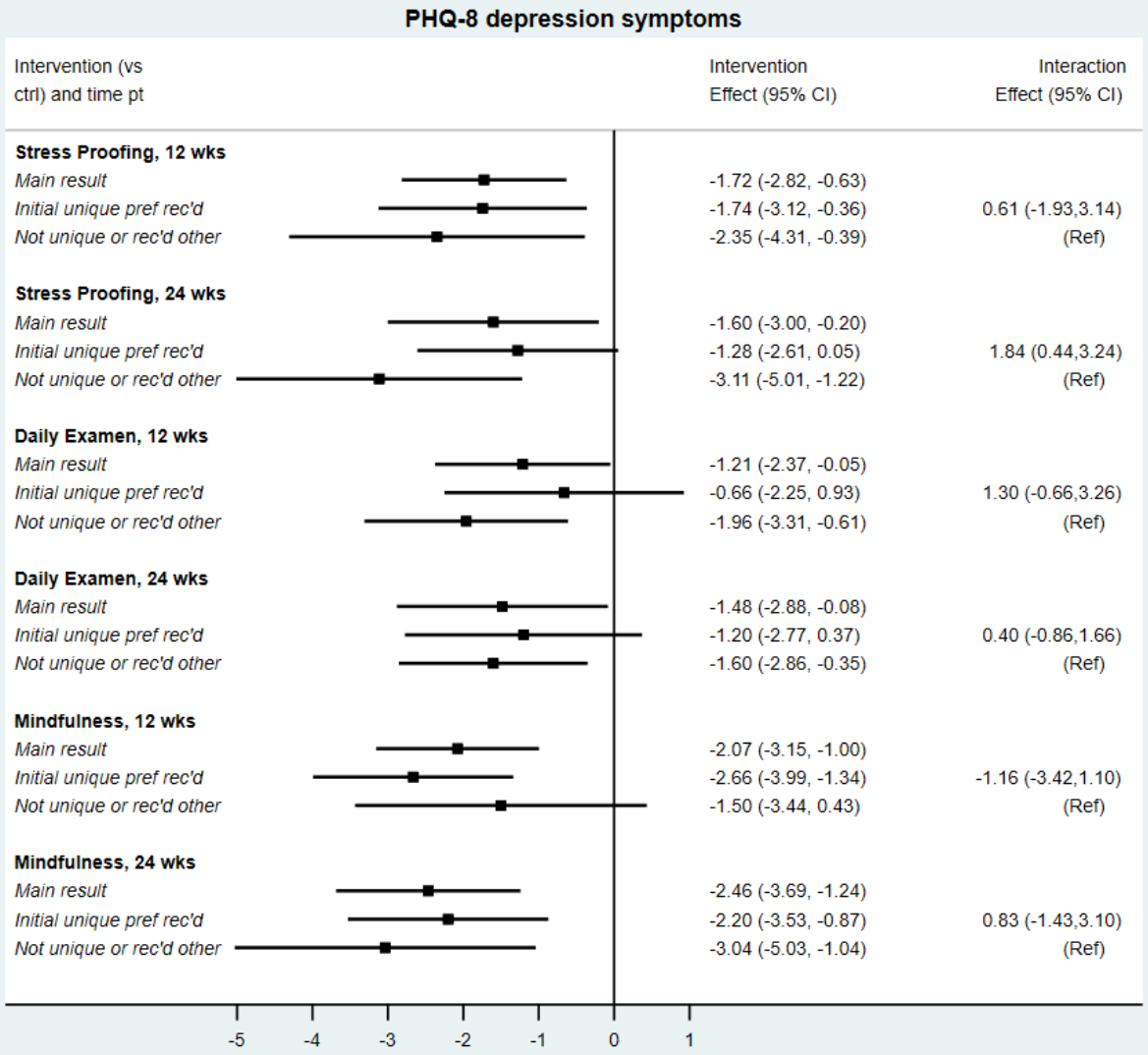
	Trial			Pooled Trial and Observational		
	No unique preference	**Had (and received) unique preference**	***p*-value**	**No unique preference**	**Had (and received) unique preference**	***p*-value**
	(N = 81)	(N = 174)		(N = 122)	(N = 232)	
Age (in years)			0.907			0.723
Mean (SD)	53.8 (11.6)	54.0 (11.0)		53.8 (11.2)	53.3 (11.6)	
Sex, n (%)			0.490			0.502
Female	41 (50.6%)	80 (46.0%)		64 (52.5%)	113 (48.7%)	
Male	40 (49.4%)	94 (54.0%)		58 (47.5%)	119 (51.3%)	
Race/Ethnicity, n (%)			0.010			0.012
White and not Latinx	69 (85.2%)	162 (93.1%)		101 (82.8%)	214 (92.2%)	
African American and not Latinx	5 (6.2%)	10 (5.7%)		10 (8.2%)	12 (5.2%)	
Asian-American/Pacific Islander, Native American, Latinx, bi/multi-racial, and other	7 (8.6%)	2 (1.1%)		11 (9.0%)	6 (2.6%)	
Marital & habitation status, n (%)			0.534			0.132
Not married, separated, or divorced	10 (12.3%)	17 (9.8%)		20 (16.4%)	25 (10.8%)	
Married or cohabitating	71 (87.7%)	157 (90.2%)		102 (83.6%)	207 (89.2%)	
Any children living at home, n (%)			0.499			0.335
No	39 (50.0%)	95 (54.6%)		61 (52.1%)	133 (57.6%)	
Yes	39 (50.0%)	79 (45.4%)		56 (47.9%)	98 (42.4%)	
Clergy appointment, n (%)			0.917			0.979
Pastoral charge	67 (82.7%)	143 (82.2%)		99 (81.1%)	188 (81.0%)	
Extension or other	14 (17.3%)	31 (17.8%)		23 (18.9%)	44 (19.0%)	
Bi-vocational, n (%)			0.918			0.633
No	78 (96.3%)	168 (96.6%)		115 (95.0%)	223 (96.1%)	
Yes	3 (3.7%)	6 (3.4%)		6 (5.0%)	9 (3.9%)	
Hours per week worked as full-time clergy			0.596			0.549
Mean (SD)	48.9 (9.9)	49.7 (10.8)		48.6 (10.6)	49.4 (10.2)	

Stress from congregation(s)/work[1], [0-3]			0.291			0.159
Mean (SD)	1.8 (0.7)	1.9 (0.7)		1.8 (0.7)	1.9 (0.7)	
Financial stress, n (%)			0.271			0.764
Not at all or slightly stressful	56 (73.7%)	116 (66.7%)		78 (67.2%)	159 (68.8%)	
Moderately, very, or extremely	20 (26.3%)	58 (33.3%)		38 (32.8%)	72 (31.2%)	
Alcoholic drink intake, n (%)			0.790			0.961
None	27 (36.5%)	59 (33.9%)		38 (33.3%)	75 (32.5%)	
Occasional drink (not every week)	16 (21.6%)	44 (25.3%)		32 (28.1%)	61 (26.4%)	
1-2 drinks	10 (13.5%)	33 (19.0%)		16 (14.0%)	40 (17.3%)	
3-6 drinks	13 (17.6%)	23 (13.2%)		15 (13.2%)	33 (14.3%)	
about a drink a day	6 (8.1%)	10 (5.7%)		10 (8.8%)	16 (6.9%)	
more than a drink a day	2 (2.7%)	5 (2.9%)		3 (2.6%)	6 (2.6%)	
Self-reported current heavy alcohol use, n (%)			0.385			0.082
No	72 (97.3%)	169 (98.8%)		110 (96.5%)	225 (99.1%)	
Yes	2 (2.7%)	2 (1.2%)		4 (3.5%)	2 (0.9%)	
Caffeinated beverage intake per day, n (%)			0.404			0.507
None	6 (8.1%)	24 (13.8%)		10 (8.8%)	29 (12.6%)	
1 cup	18 (24.3%)	41 (23.6%)		29 (25.4%)	56 (24.2%)	
2-3 cups	36 (48.6%)	82 (47.1%)		53 (46.5%)	113 (48.9%)	
4-5 cups	13 (17.6%)	20 (11.5%)		19 (16.7%)	25 (10.8%)	
6 or more cups	1 (1.4%)	7 (4.0%)		3 (2.6%)	8 (3.5%)	
Metabolic equivalents (METs) per week			0.015			0.001
Mean (SD)	48.1 (71.8)	75.5 (87.8)		43.4 (67.4)	74.0 (87.5)	
Body Mass Index (BMI)			0.001			0.010
Mean (SD)	32.8 (8.3)	29.8 (6.1)		32.0 (7.8)	30.0 (6.3)	
Obesity, n (%)			0.020			0.117
Not obese (BMI <30)	33 (40.7%)	98 (56.3%)		54 (44.3%)	123 (53.0%)	
Obese (BMI 30+)	48 (59.3%)	76 (43.7%)		68 (55.7%)	109 (47.0%)	
Hypertension, n (%)			0.263			0.133
No (including missing)	57 (70.4%)	110 (63.2%)		88 (72.1%)	149 (64.2%)	
Yes, current or history	24 (29.6%)	64 (36.8%)		34 (27.9%)	83 (35.8%)	
Diabetes, n (%)			0.119			0.056

No (including missing)	67 (82.7%)	156 (89.7%)		102 (83.6%)	210 (90.5%)	
Yes, current or history	14 (17.3%)	18 (10.3%)		20 (16.4%)	22 (9.5%)	
PHQ-8 depression symptoms sum score, [0-24]			0.991			0.555
Mean (SD)	5.4 (4.8)	5.4 (4.5)		5.6 (5.0)	5.3 (4.5)	
Depression screens, n (%)			0.783			0.542
Negative (PHQ-8 <10)	68 (84.0%)	142 (82.6%)		97 (79.5%)	189 (82.2%)	
Positive (PHQ-8 10+)	13 (16.0%)	30 (17.4%)		25 (20.5%)	41 (17.8%)	
C-SOSI stress symptoms mean score, [0-4]			0.672			0.813
Mean (SD)	1.0 (0.6)	1.0 (0.5)		1.0 (0.6)	1.0 (0.5)	
GAD-7 anxiety sum score, [0-21]			0.682			0.795
Mean (SD)	4.6 (4.8)	4.8 (4.3)		4.8 (4.8)	4.6 (4.2)	
HRV MESOR			0.929			0.998
Mean (SD)	24.6 (13.1)	24.4 (17.3)		25.1 (14.0)	25.1 (17.5)	

[1] From November 2019 to registration for Trial participants; from February 2020 to registration for Observational participants

Supplemental Figure B5. Subgroup analysis of heterogeneity of treatment effects on depression by preference type



Supplemental Table B12. Intra-class correlation coefficients for clustering due to group delivery of intervention estimated using mixed effects regression for trial period participants

	Unadjusted	**Adjusted**
Survey outcomes		
C-SOSI	0.061	0.000
GAD-7	0.023	0.000
PHQ-8	0.035	0.007
HRV outcomes		
MESOR	0.000	0.000
Amplitude	0.000	0.000

Appendix C: Statistical Analysis Plan (SAP)

Title	Selah Trial: A randomized preference trial of 3 stress reduction interventions
CRU/Department/Division/Center	Clergy Health Initiative
IRB Number	2019-0238
Investigators	
Primary Investigator	Rae Jean Proeschold-Bell
Collaborative Lead	David Eagle, Josh Rash
Co-authors (if known)	Logan Tice, Jessie Larkins, Jia Yao, Timothy Kim, Liz Wallack
Analysis Biostatistician(s)	Jia Yao
Biostatistics Supervisor	Liz Turner
Lead Biostatistician	Alyssa Platt
Subject Matter Expert	Josh Rash (HRV and stress reduction)
Original Creation Date	Feb 23, 2021
Version Date	Jun 28, 2021
Project Folder Location	Box\CHI Writing Group\Selah Trial Outcomes_RJ
Project Goal(s)	Manuscript
Submission Deadline(s)	April 2023
Effort Estimate (optional)	

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.
☒ If substantial additional analysis is necessary or the aims of the project change, a new SAP will need to be developed.
☒ Publications resulting from this SAP are supported in part by the Duke CTSA and must cite grant number UL1TR002553 and be submitted to PubMed Central.
☒ I have reviewed the SAP and understand that any changes must be documented.

Acknowledged by: Click or tap here to enter text.
Date: Click or tap to enter a date.

Activity Log

- (3/16/2021): Preliminary draft SAP completed
- (2/25/2022): Preliminary analysis completed
- (3/1/2022): Study data unblinded, and unblinded analysis performed
- (2/25/2022): Due to baseline imbalances between study arms, propensity score outcome switched from preferred intervention to assigned intervention

	• (4/14/2022): With satisfactory balance via propensity score adjustment, longitudinal analysis method switched from LDA to cLDA • (6/1/2022): Primary estimands changed from ITT to AT • (6/8/2022): From study team discussion of goals of inference, multiple comparisons adjustment switches from disjunction testing for 3 intervention types (separately by outcome) to individual testing for each intervention type adjusting for two outcomes (HRV and C-SOSI) • (11/23/2022): Depression scale added as an exploratory outcome

Acronyms		
	AT	As-treated
	C-SOSI	Calgary Symptoms of Stress Inventory
	DE	Daily Examen
	ECG	Electrocardiogram
	GAD	Generalized Anxiety Disorder
	HRV	Heart Rate Variability
	ICC	Intraclass correlation coefficient
	LTI	Leisure Time Index
	MESOR	Midline Estimating Statistic of Rhythm
	MBSR	Mindfulness-Based Stress Reduction
	PHQ	Patient Health Questionnaire
	RMSSD	Root Mean Square Successive Difference
	SP	Stress Proofing
	UMC	United Methodist Church

1 Study Overview

1.1 Background/Introduction:

Certain populations may be particularly susceptible to the adverse effects of chronic stress, particularly chronic work-related stress. One such population is that of clergy doing ministry work. Clergy exhibit high prevalence rates of chronic disease, including diabetes, hypertension, asthma, and joint-related disease, as well as obesity.[1] Further, studies also indicate above-average rates of depression.[2] While no studies directly compare rates of clinical anxiety among clergy versus non-clergy, anxiety rates among clergy are elevated. The high prevalence of physical and mental health issues among clergy may be due in part to stressors from the unique nature of clergy work.[3]

Researchers have developed numerous approaches to manage stress, such as cognitive-behavioral therapy, mindfulness, and relaxation.[4] Not all stress-management interventions are equally effective. Rather, stress-reducing activities are viewed as skills that require regular practice.[5] As such, the most effective interventions are those that individuals are willing and motivated to practice (i.e., patient-preference is an important aspect of evidence-based practice).[6] The current study builds on a pilot study that the investigators conducted with clergy to evaluate the feasibility and acceptability of four potentially stress-reducing interventions while taking participant preference into account. Three of those stress-reducing interventions showed trends of self-reported stress reduction in terms of reduced stress symptoms and/or reactivity to stress. In the current study, the investigators test those three interventions: Mindfulness-Based Stress Reduction, the Daily Examen, and a set of stress inoculation and breathing exercises called Stress Proofing.

This trial tries to answer 1) among three stress reduction interventions, which are acceptable to clergy and 2) whether and for how much each intervention leads to reductions in survey-based stress symptom measures and improvements in heart rate variability.[1]

ClinicalTrials.gov trial registration number: NCT04625777.

1.2 Study Aims

The aim of this study is to evaluate the effectiveness and acceptability of three stress reduction interventions: Mindfulness-Based Stress Reduction, the Daily Examen, and Stress Proofing, on self-reported stress symptoms and heart rate variability.

1.2.1 **Aim 1**. Descriptive analysis of study participation (e.g. session attendance, acceptability data)

1.2.2 **Aim 2**. Main trial analysis

1.3 Research Objectives and Study Hypotheses

1.3.1 Objective 1: To compare the impact of, separately, MBSR, the Daily Examen, and Stress Proofing with a waitlist control condition, on stress outcomes at 12 weeks.

1.3.1.1 *Hypothesis 1. When randomly assigned to waitlist vs non-waitlist, MBSR, the Daily Examen, and/or Stress Proofing will be superior to the waitlist control on stress symptoms (measured by C-SOSI score) at 12 weeks*

1.3.1.2 *Hypothesis 2. When randomly assigned to waitlist vs non-waitlist, MBSR, the Daily Examen, and/or Stress Proofing will be superior to the waitlist control on heart rate variability (HRV) at 12 weeks (co-primary outcomes).*

1.3.2 Objective 2: To compare the impact of, separately, MBSR, the Daily Examen, and Stress Proofing with a waitlist control condition, on anxiety symptoms at 12 weeks.

1.3.2.1 *Hypothesis 3. When randomly assigned to waitlist vs non-waitlist, MBSR, the Daily Examen, and/or Stress Proofing will be superior to the waitlist control on anxiety symptoms (measured by GAD-7 score) at 12 weeks (secondary outcome).*

1.3.3 Objective 3: To compare the impact of, separately, MBSR, the Daily Examen, and Stress Proofing with a waitlist control condition, on stress and anxiety symptoms at 24 weeks.

1.3.3.1 *Hypothesis 4. When randomly assigned to waitlist vs non-waitlist, MBSR, the Daily Examen, and/or Stress Proofing will be superior to the waitlist control on stress symptoms (measured by C-SOSI score) at 24 weeks (exploratory outcomes).*

1.3.3.2 *Hypothesis 5. When randomly assigned to waitlist vs non-waitlist, MBSR, the Daily Examen, and/or Stress Proofing will be superior to the waitlist control on anxiety symptoms (measured by GAD-7 score) at 24 weeks (exploratory outcomes).*

1.3.4 Objective 4: To determine whether having a preference, and that preference being honored, was associated with better outcomes on stress symptoms, HRV, and anxiety symptoms.

1.3.4.1 *Hypothesis 6. Participants who had an initially stated unique preference and received that intervention (i.e. MBSR, the Daily Examen, and Stress Proofing combined) will experience larger between-arm (waitlist vs non-waitlist) differences in improvements on stress symptoms (measured by C-SOSI score) at 12 weeks when compared to participants that did not have a unique preference or did not receive their initially uniquely preferred intervention (exploratory outcome).*

1.3.4.2 *Hypothesis 7. Participants who had an initially stated unique preference and received that intervention (i.e. MBSR, the Daily Examen, and Stress Proofing combined) will experience larger between-arm (waitlist vs*

non-waitlist) differences in improvements on HRV at 12 weeks when compared to participants that did not have a unique preference or did not receive their initially uniquely preferred intervention (exploratory outcome).

1.3.4.3 *Hypothesis 8. Participants who had an initially stated unique preference and received that intervention (i.e. MBSR, the Daily Examen, and Stress Proofing combined) will experience larger between-arm (waitlist vs non-waitlist) differences in improvements on anxiety symptoms at 12 weeks when compared to participants that did not have a unique preference or did not receive their initially uniquely preferred intervention (exploratory outcome).*

2 Methods

2.1 Sample Size

A non-randomized Selah pilot study was conducted prior to the full trial, using participant preference design with a control group. The pilot study was used to evaluate the feasibility, acceptability, and range of effect sizes in outcomes for four potentially stress-reducing interventions: stress inoculation training (Stress Proofing), mindfulness-based stress reduction (MBSR), the Daily Examen, and Centering Prayer[2]. 78 clergy participated in an intervention and 7 provided data as a control group. No randomization was performed, therefore baseline stress levels varied widely between treatment groups.

2.1.1 Table 1. Baseline and follow-up means for stress symptoms observed during the Selah Pilot study

		Mean (SD)	
	Sample size at baseline	0 weeks	12-weeks[1]
C-SOSI stress symptoms, scale range 0-4	77		
MBSR	12	1.02 (0.46)	0.70 (0.58)
Stress Proofing	29	0.90 (0.50)	0.55 (0.36)
Daily Examen	17	0.65 (0.42)	0.51 (0.38)
Centering Prayer	12	0.40 (0.27)	0.39 (0.27)
Control	7	0.56 (0.38)	0.59 (0.38)

[1]Closer to 8-17 weeks for control participants (median=12 weeks)

To estimate probable baseline and follow-up scores for the full trial (which would be more likely to achieve baseline balance in stress scores due to the randomized waitlist design), a baseline average C-SOSI score was calculated across all groups (excluding those that participated in Centering Prayer). The average baseline C-SOSI score was 0.92 (SD=0.46). Estimates for probable C-SOSI scores at the 12-week follow-up time point used observed 12-week scores of 0.7 (SD=0.58) for MBSR, 0.55 (SD=0.36) for Stress Proofing, and 0.51 (SD=0.38) for the Daily Examen.

Given an alpha of 0.0167 (based on a Bonferroni correction to hypothesis tests for the effects of 3 interventions on C-SOSI scores), a per-arm sample size of 195 for MBSR, 40 for Daily Examen, and 47 for Stress Proofing will yield 80% power to detect a between-arm difference in means at 12-weeks of 0.22 for MBSR, 0.37 for Stress Proofing, and 0.41 for Daily Examen (compared to waitlist control) for a two-sample t-test with unequal variances, allowing for loss-to-

[2] Note that the Centering Prayer intervention was removed as an intervention option for the full trial due to low enrollment and lack of preliminary evidence of potential efficacy

follow-up of 20% and a design effect of 1.3 (corresponding to an ICC of 0.027 and average cluster size of 12)[3] to account for clustering caused by the group-based intervention delivery.

Previous literature recommends that medium effect size (difference in means/standard deviation) for HRV be defined as 0.50. A per-arm sample size of 140 will yield 80% power to detect an effect size of 0.50 for a two-sample t-test with an alpha of 0.0167, allowing for loss-to-follow-up of 20% and a design effect of 1.3 to account for group-based intervention delivery. All power analyses were performed using PASS 2021 software.

2.2 Randomization

All participants recruited before March 1, 2020 were randomly assigned to non-waitlist vs waitlist arms. Data on participants' preference on the three interventions were collected before randomization. Table 2 reports assignment and randomization allocations for each preference scenario. The analysis statistician wrote the randomization codes in Stata version 16[4]. A secondary recruitment period was initiated after March 1, 2020 that did not include random assignment; data from those participants will be included in a secondary, observational analysis. The analysis statistician will remain blinded to intervention allocation and waitlist randomized assignment until the analysis plan has been finalized and initial blind reviews have been performed.

Importantly, baseline data collection occurred AFTER randomization, so it is possible for dropout to have occurred prior to baseline data collection.

2.2.1 Table 2. Study assignment approach for each preference and enrolment date scenario

Preference Scenario	**Assignment Approach**
Enrolled prior to March 1, 2020	
No preference between interventions	Randomly assigned to one of the four study arms: three interventions without waitlist and the waitlist arm, with a 1:1:1:1 ratio
Preferred two interventions equally and over the third intervention	Randomly assigned to one of the two interventions with a 1:1 ratio, and then a fraction was randomly assigned to the waitlist arm, with a 3:1 non-waitlist vs waitlist ratio for MBSR and Stress Proofing and a 5:4 non-waitlist vs waitlist ratio for the Daily Examen (DE was preferred by more participants and this allowed the same number of participants to be randomized into each study arm)
Preferred one intervention among the three	Assigned to their preferred intervention and combined with participants with two top preferences who had been randomized to that intervention, and then randomly assigned to non-waitlist vs waitlist arms, with a 3:1 non-waitlist vs waitlist ratio for MBSR and Stress Proofing and a 5:4 non-waitlist vs waitlist ratio for the Daily Examen
Any of the above scenarios and are part of a married (or cohabitating) couple who both meet study criteria and enrolled	To avoid spillover effects, each couple was treated as if they were one person, i.e., assigning both spouses to the same intervention and randomizing them into a non-waitlist vs waitlist arm. When a couple had different preferences, one preference was randomly chosen as the couple's preference.
Seven clergy with an established meeting group (a covenant group)	The seven clergy jointly chose a single preferred intervention and were randomized together to the non-waitlist vs waitlist arm.
Enrolled after March 1, 2020	
All enrollees after March 1, 2020 with no preference, two equal	Participants answered treatment preference survey items but regardless of their answers, self-selected the intervention with intervention dates they most

[3] Calculated as $Design\ Effect = 1 + \delta(n-1)$ where δ is ICC and n is average cluster size. Note that this conservatively assumes there is clustering throughout the sample, however, we expect only partial clustering due to group treatment delivery.
[4] The randomization codes are hosted in Box\PROJECT_DGHI_SELAH_rproesh\randomization\programs

preferences, or one preference, and whether a clergy couple or not	wanted from the full list of workshop options. None were randomized to the non-waitlist vs waitlist arms; they all were assigned to non-waitlist.

2.3 Blinding and Masking

It is not possible to blind study participants to treatment selection. Blinding of analysts is complicated by the structure and design of the data (e.g. the waitlist participants can have up to 6 study observations, immediate intervention only up to 3). For this reason, blinding of analysts will occur in two ways:

1. Assignment of ambiguous labels to study arm assignment variables by study team member other than primary analyst and supervising statistician
2. Within participant, random permutation of survey timepoint

Ambiguous labels will be applied so that the analyst cannot be influenced by knowledge that a participant received a particular intervention type. Permutation of time points is intended to nullify the treatment effect in order to execute the analysis plan without trying to "optimize" the magnitude of the treatment effect (since an ambiguous label cannot be used with the waitlist arm). Once a preliminary study analysis plan has been executed, all data will be unblinded.

2.4 Study Population

2.4.1 Inclusion Criteria

- Clergy under appointment (including retired but still under appointment) in the 2019-2020 or 2020-2021 appointment cycle in the Western NC Conference or the NC Conference of the UMC as an elder, deacon, local pastor, or extension minister (broadly defined, including district superintendent and bishop)
- 18 years of age or older in the 2019-2020 or 2020-2021 appointment cycle
- Willing to participate in the survey and (if health allows) HRV data collection and commit to completing their assigned stress management intervention

2.4.2 Exclusion Criteria

- Participants with underlying medical conditions which could seriously impact the integrity of their HRV data were excluded from HRV data collection, including
    - A diagnosis of COVID-19;
    - A diagnosis of tachycardia;
    - Being pregnant or becoming pregnant during the course of data collection;
    - Having a pacemaker, or internal defibrillator;
    - Documentation of other cardiovascular-related chronic or acute morbidities, including
        - Irregular heartbeat (arrhythmia), taking medication for the treatment of irregular heart rate, and/or have ever had surgery to correct an abnormal or irregular heart rate
        - Have ever had a heart attack or stroke

2.5 Study Procedure

Prior to March 1, 2020 and the COVID-19 pandemic, intervention assignment occurred using a partially randomized preference design[7] (See Appendix Figure 1). Participants read descriptions of the three stress management interventions and provided preference ratings using the Treatment Acceptability and Preferences Scale[8]. In addition, participants were asked whether they preferred any of the three interventions or if the interventions were equally appealing to them (i.e. no preference). Those with a preference were asked whether they had one or two equal first choice(s).

Participants who preferred one intervention were assigned to their preferred intervention and randomized to a non-waitlist vs waitlist condition; the waitlist condition participants provide control data prior to intervention receipt. Upon completing the waitlist condition, participants were allowed to update their preference and receive their

currently preferred intervention while providing data. Participants who stated no preference (or equally preferred two interventions over the third) were randomized to receive one of three (or, if applicable, two) interventions in a non-waitlist condition or to the waitlist condition.

After March 1, 2020 we continued to recruit participants and assigned them to their preferred intervention in a non-waitlist condition. These additional recruits will provide observational data. Note that unlike with the participants recruited prior to March 1, 2020, there was no randomization process to assign participants with no preference or tied preference to an intervention. The participants had to choose one of the 3 interventions (similar to the structure of the Selah pilot study).

2.5.1 Study Arms: Non-waitlist interventions and waitlist

Interventions will be delivered in small groups of 10-25 participants. Survey and text message data will be collected online; HRV data will be collected by the participants in their respective personal settings.

2.5.1.1 Mindfulness-Based Stress Reduction (MBSR)

The specific course for this study is a synchronous web-based video platform course with certified instructors from Duke Integrative Medicine and content based on Jon Kabat-Zinn's model.[9] It includes exercises in awareness of breath, body scans, walking meditation, "choiceless" open awareness, Loving Kindness Meditation, and bringing awareness to the present moment. The course consists of 8 weekly sessions confined to study participants held via video conference with meditation instruction, periods of guided practice, and group discussion. Participants are also offered a 4-hour online “Day of Mindfulness” which additionally includes community members not enrolled in the study.

From baseline to 24 weeks, participants receive daily text messages that ask the number of minutes (which could include 0) they practiced MBSR yesterday.

2.5.1.2 The Daily Examen Prayer Practice (DE)

The Daily Examen guides the person through a five-step prayer.

1. Become aware of God’s presence.
2. Review the events of the past 24 hours, recalling 2-3 things for which you are grateful.
3. Review the events of the past 24 hours, guided by the Holy Spirit, noticing where you experienced God’s presence.
4. Review what stands out and pay attention to what emotions arise. With the guidance of the Holy Spirit, pray through these emotions, noticing which are drawing you closer to God or pulling you away from God.
5. Look forward to the next 24 hours. What is one thing you should do? Where do you need God’s assistance?

The Daily Examen is designed to help the supplicant reflect on positive emotions, move past negative emotions, and align their work with God's work.[10] Instructors developed workshop content that included three occasions of practicing the Daily Examen, lecture instruction, and small group discussion. Topics covered included a history of Ignatian spirituality, the emotions and the spiritual life, and practicalities of developing a daily prayer practice. We asked participants to conduct the Daily Examen as a daily practice, requiring a 10-15-minute commitment for 6 months following their workshop. Participants were also offered the opportunity to meet with their instructors in an online small group format two and also six weeks following their workshop to address any issues arising from their practice.

From baseline to 24 weeks, participants receive daily text messages that ask if they practiced the Examen yesterday (yes/no).

2.5.1.3 Stress Proofing inoculation combination (SP)

Stress Proofing is a set of stress reduction skills with aspects of stress inoculation training.[11] The founder designed and led a weekly, synchronous, web-based workshop for four weeks. Consistent with stress inoculation training, the workshop began with education on the stress response and physical awareness of one's own stress response.[12] The training focuses on physical activities to undo the stress response including walking with diaphragmatic breathing, triangle and rectangle breathing, tension control, stretching, and massage. Stress inoculation training is discussed and participants are encouraged during periods of less stress to allow themselves a degree of physical discomfort to learn to tolerate discomfort in the future. The training recommends a variety of beneficial lifestyle practices, including prioritizing nutrition and sleep and disengaging from technology for several hours before sleep. The daily practice plan emphasized stress awareness and diaphragmatic breathing, with encouragement to try the lifestyle adjustments for a week at a time.

From baseline to 24 weeks, participants receive daily text messages that ask if they did 0, 1, or 2 resets yesterday.

2.5.1.4 Waitlist condition

Participants allocated to the waitlist arm will undergo a 24-week waiting period, during which they will respond to surveys (0, 12, and 24 weeks). Those who are eligible will also provide a 48-hour continuous ambulatory heart rate recording at 0 and 12 weeks (HRV)[5]. Upon completing the waiting period, participants will be informed that they can update their intervention preference and receive that preferred intervention while providing intervention data. Once waitlist participants reach the point at which they will begin to receive an intervention, they then provide intervention data at baseline (immediately prior to start of intervention), 12, and 24 weeks similar to those that were randomized to the immediate receipt of intervention.

Surveys for waitlist participants were staggered in time in groups of 20, and were agnostic to trial preference/assignment.

Waitlist in as-treated analysis includes participants who were assigned to a non-waitlist intervention but did not participate in the intervention and only provided survey/HRV data. The variable that indicates as-treated arms was generated in REDCap using the number of intervention workshop sessions attended i.e. as-treated waitlist consists of participants who attended 0 intervention sessions.

2.6 Data Acquisition

Study design	Interventional; preference-based randomized waitlist trial
Data source/ how the data were collected	Registration data: collected through and stored in Qualtrics. There were multiple registration periods due to COVID-related delays. Participation data: stored in REDCap (Project name: "Selah Trial"; PID: 8246) Survey data: stored in REDCap **There are two separate datasets within REDCap**:

[5] Note that if randomized to the waitlist condition, a participant may have provided HRV data for up to 4 time points (0- and 12-weeks during waitlist; 0- and 12-weeks during intervention period)

	• Selah_A (REDCap Project name: "Selah Trial"; PID: 8246) dataset contains intervention participants and waitlist control participants with the first three time points (baseline, 12 weeks and 24 weeks)
	  ○ **For intervention participants**[1] - email REDCap surveys at baseline (immediately prior to intervention), 12, and 24 weeks
	  ○ **For waitlist control participants** – email REDCap surveys at 0 weeks (corresponding to timing of non-waitlist interventions baseline survey), 12, and 24 weeks[2]
	• Selah_B (REDCap Project name: "Selah Trial_b"; PID: 10450) dataset only contains control participants or "post-waitlist participants" with their last three time points - email REDCap surveys at post-waitlist baseline (immediately prior to intervention), 12, and 24 weeks
	• Selah_A and Selah_B datasets will be blinded and merged to create one complete master survey dataset
	HRV data:
	Collected using eMotion Faros 180 ambulatory heart rate recording device (Bittium) connected by electrode leads to two pre-gelled (Ag/AgCl) disposable Ambu BlueSensor wet-gel ECG electrodes attached beneath the right clavicle and left ribcage. Worn for 48 hours at baseline and 12 weeks.
	
	Text message data:
	Daily texts will be sent to participants from intervention baseline (i.e. when intervention began; NOT at waitlist 0 weeks) to 24 weeks after intervention started, responses will be transmitted automatically to Prompt.

Contact information for team member responsible for data collection/ acquisition	Logan.tice@duke.edu (Logan Tice) for registration, participation, survey and HRV data jarash@mun.ca (Josh Rash) for questions about HRV data cleaning and processing beth.stringfield@duke.edu (Beth Stringfield) for text message data
Date or version (if downloaded, provide date)	Registration data: latest Aug 16, 2021 Participation & Survey data: latest Oct 2, 2021 HRV data: latest Jan 8, 2022 Text message data: latest Nov 11, 2021
Data transfer method and date	Registration data – downloaded from Qualtrics and saved in a secure Duke Box Participation & Survey data – downloaded from REDCap and saved to a secure Duke Box HRV data - downloaded from heart rate monitor and saved to a secure Duke Box Text message data – downloaded from Prompt and transferred to a secure Duke Box
Where dataset is stored	Registration data: • **For early enrollees (before Mar 1, 2020) -** Box\PROJECT_DGHI_SELAH_rproesh\randomization\data\ Selah Registration 2.24.2020 IDs only.xlsx • **For late enrollees (after Mar 1, 2020) -** Box\PROJECT_DGHI_SELAH_rproesch\Data collection\[NOT YET BLINDED] Data Cleaning & Organization\Preference Data\Spirited Life_ Selah Initial Registration (Re-Open)_August 16, 2021_12.38.xlsx Box\PROJECT_DGHI_SELAH_rproesch\Data collection\[NOT YET BLINDED] Data Cleaning & Organization\Preference Data\ Spirited Life_ Selah Initial Registration and Workshop Registration 2021_August 16, 2021_12.43.xlsx ***There are two excel files for late enrollees. Survey and participation data: Box\PROJECT_DGHI_SELAH_rproesh\Data collection\[not yet blinded] Data Cleaning & Organization

	HRV data: Box\PROJECT_DGHI_SELAH_rproesch\HRV Data Trial\[UNBLINDED} Finalized HRV from Josh & Liz Text message data: [3] • Original data are housed in Prompt – https://prompt-production.herokuapp.com/users/sign_in • Outside of Prompt, there are downloads at Box /PROJECT_DGHI_SELAH_rproesh/Data Collection/Text Messaging/Data Backups • Also, at the end of each cohort, that end cohort's data will be separated out - Box/Selah/Data Collection/Text Messaging/Data Tracking & Reports (not blinded)/Cohort close-out reports --- all cohorts starting with 7_DE 4. The naming convention is xxxx (e.g., 7), intervention name (e.g., DE for Daily Examen), cohort (e.g., 4)
Where the data dictionaries are stored	For the registration data: Box\PROJECT_DGHI_SELAH_rproesch\Data collection\Registration survey\codebooks For the Survey data: Box\PROJECT_DGHI_SELAH_rproesch\Data collection\trial survey data\Codebook For HRV data: Box\PROJECT_DGHI_SELAH_rproesch\Data release\HRV data\Data dictionary Other data dictionaries: Box\PROJECT_DGHI_SELAH_rproesh\Data dictionaries

[1]*Participants post March 1, 2020 have survey data collection that mirrors intervention (non-waitlist) participants*

[2]*Once waitlist control period ends, additional survey data is the same as for "For intervention participants"*

3 Outcomes, Exposures, and Additional Variables of Interest

3.1 Primary Outcome(s)

Outcome	Description	Variable Name/Source	Specifications
Self-reported stress symptoms	Modified version of the Calgary Symptoms of Stress Inventory.[13] (Scale range: 0-4), based on average score on 4 item scale.	Variable name: csosi Source: Box\PROJECT_DGHI_SELAH_rproesh\Data collection\trial survey data\3_data with derived vars\survey_wdrv_latest.dta	Use in continuous form
Heart Rate Variability: Midline Estimating Statistic of Rhythm	48-hour ambulatory heart rate data using ECG devices.[6] Heart rate variability indexed using the time-domain	Variable names: mesor; amplitude Source: Box\PROJECT_DGHI_SELAH_rproesch\HRV Data Trial\[BLINDED] HRV data\Selah HRV Datafile seg removed_Transposed_Winsorized_participants without data removed_blinded.dta	Two individual-level cosine function parameters will be estimated to quantify the circadian variability parameters: i) MESOR, defined as the rhythm adjusted 24-hour mean,

[6] *Participants will be fitted with an eMotion Faros 180 ambulatory heart rate recording device (Bittium) connected by electrode leads to two pre-gelled (Ag/AgCl) disposable Ambu BlueSensor wet-gel ECG electrodes attached beneath the right clavicle and left ribcage. The 48-hour ECG recording will be imported to Kubios HRV Premium V3.4.1 software, partitioned into 5-minute segments, visually inspected to allow for manual correction of ectopic beats, detrended, and subject to Kubios' automatic artefact correction algorithm.*

(MESOR) and amplitude	metric Root Mean Square of Successive RR Differences (RMSSD).		and ii) amplitude, defined as the distance between MESOR and the maximum of the cosine curve (i.e. half the extent of rhythmic change in a cycle).

3.2 Secondary Outcome(s)

Outcome	Description	Variable Name/Source	Specifications
Self-reported anxiety symptoms	We will use continuous scores of the seven-item Generalized Anxiety Disorder-7 (GAD-7) measure. (Scale range: 0-21), based on sum score	Name: gads Source: Box\ PROJECT_DGHI_SELAH_rproesh\ Data collection\trial survey data\3_data with derived vars\survey_wdrv_latest.dta	Use in continuous form. Additionally, dichotomized right below 8: - Anxiety screens positive (GAD7 sum score 8-21) - Anxiety screens negative (GAD7 sum score 0-7)

3.3 Additional Variables of Interest

Source of all these variables: Box\PROJECT_DGHI_SELAH_rproesh\Data collection\trial survey data\3_data with derived vars\survey_wdrv_latest.dta

Variable	Description	Variable Name	Specifications
Sex	Response options: Male; Female. Questionnaire actually asked gender in the wording	d13	As is
Age	Participants were asked to write in number of years old	d16	Continuous, as is
Race	Response options (allowing multiple choices): White; African American; Asian-American/Pacific Islander; American Indian or Alaskan Native; Other (specify)	d15	Collapsed not mutually exclusive categories for analysis: 1. White 2. Non-White (note: would prefer more nuance here but categories collapse to only 2 because of sparsity in races other than White)
Ethnicity	Response options: Latinx; Not Latinx	d14	Exclude from analysis due to sparsity

Marital (and cohabitation) status	Response options to marital status: Married, Not married, Married but separated and/or in the process of divorcing; Not-married participants were asked whether or not they were cohabitating with significant other.	d21, d21_2	Dichotomize as: - Married or cohabitating - Not married, or married but separated/divorcing
Child(ren) living at home	Response options: Yes; No	d22	As is
Bi-vocational status	Whether participant has another job in addition to serving as clergy Response options: Yes; No	ax7	As is
Clergy appointment	Response options: Pastoral charge appointment; Extension or other kind of ministry appointment	ax1	As is
Number of hours worked per week (as UMC clergy)	Participant was asked to write in numeric values in number of hours per week	ax9	Continuous (as is)
Number of congregants	Participant was asked to report the total number of people (adults and children) that attend participant's worship services in an average week	d12	Categorized as: - 1-149 people in worship per week; - 150+ people per week; - Participant is not appointed to church
Financial stress	Response options: Not at all stressful; Slightly; Moderately; Very; Extremely	d20	Dichotomize for analysis: - Not at all/slightly - Moderately/very/extremely
Depressive symptoms	We will use continuous scores of the eight-item Patient Health Questionnaire (PHQ-8) which asks about the frequency of specific depressive symptoms experienced within the past two weeks. Scale range: 0-24	b3a-b3h	Dichotomized as: - Depression screens positive (PHQ8 sum score 10-24) - Depression screens negative (PHQ8 sum score 0-9)
Self-reported mental health conditions	Conditions: Depression, Anxiety, Schizophrenia, Bipolar disorder; Eating disorder, Heavy alcohol use, Drug dependency	d9l_21, d9l_22, d9l_24, d9l_25; d9l_23, d9l_26, d9l_27	Dichotomize as: - Yes, currently or history - No or don't know

	Response options: No; Yes, currently; Yes, I have a history but not currently; Don't know.		* For heavy alcohol use, currently yes vs currently no is reported in sample description;
Self-reported physical health conditions	Conditions: High blood pressure, Chest pains or angina, Irregular heart beat, Abnormal ECG during the last 12 months, Heart attack during the last 12 months, Stroke during the last 12 months, Any other heart or circulatory problem; Diabetes, Asthma, Respiratory disorder, Cancer, Gastrointestinal disorder, Ulcer, Allergies, Arthritis or joint disorder, Reproductive system disease or disorder, Anemia, High cholesterol, Thyroid disease, Pain related disorder, Bariatric surgery during the last 12 months Response options: No; Yes, currently; Yes, I have a history but not currently; Don't know.	d9l_1, d9l_2, d9l_3, d9l_4, d9l_5, d9l_6, d9l_7; d9l_8, d9l_9, d9l_10, d9l_11, d9l_12, d9l_13, d9l_14, d9l_15, d9l_16, d9l_17, d9l_18, d9l_19, d9l_28, d9l_20	Dichotomize as: - Yes, currently or history - No or don't know
Body mass index	We will use self-reported weight (in pounds) and height (in feet and inches) to assess body mass index. BMI=Weight (lbs)/[height (ft)]^2	d8a, d8a_2, d8b	Use as is continuous Dichotomize as: - Obese (BMI>=30) - Not obese (BMI <30)
Physical activity levels	Physical activity levels will be measured using the Godin-Shephard Leisure-Time Physical Activity Questionnaire, a self-report measure of how often in the past seven days and for how many minutes per time one has engaged in physical activity, measured separately for strenuous, moderate, and mild exercise.	d10a, d10b, d10c	Total metabolic equivalents (METS) per week = [total minutes of strenuous exercise *9 + total minutes of moderate exercise *5 + (total minutes of light exercise *3] /15.[1] METs may be more precise than Leisure Time Index (LTI) as it utilizes all available information.
Average daily caffeine intake[2]	Response options (in number of cups): 0; 1; 2-3; 4-5; 6+	d9m	As is
Average weekly alcohol consumption[2]	Response options (in number of drinks per week): 0; occasional (not every week); 1-2; 3-6; 1 per day; more than 1 per day	d9n	As is

Preference for online vs in-person intervention	Response options: Strongly prefer in-person workshop; Slightly prefer in-person workshop; No preference; Slightly prefer online workshop; Strongly prefer online workshop	d25	Dichotomize as: - Prefer in-person workshop; - Prefer online workshop/no preference
Preference of treatment	Preferred MBSR; preferred SP; preferred DE; preferred 2-3 treatments		As is
Preference vs service			Dichotomize as: - Participant had a single preference and received the service that they preferred - Participant didn't have a single preference, or that they didn't receive the single treatment that they preferred

[1] https://cebp.aacrjournals.org/content/cebp/16/3/430.full.pdf

[2]Alcohol consumption and caffeine intake tend to have linear effects on HRV, and cutpoints may not be as pertinent.

4 Statistical Analysis Plan

4.1 Analysis Plan for Aim 1

Descriptive statistics will be used to summarize participation, acceptability, and compliance with interventions with summaries pooled and by study arm. Refer to "Box\CHI Writing Group\Selah Trial Outcomes_Rae Jean\sample_description\Table1a_trial.docx" for a sample output table. In order to assess generalizability, a comparison population, the 2019 Panel participants (appointed and actively serving; excluding fully retired, on leave, disabled, and those who had left UMC pastoral ministry) will additionally be described.

Text message data, taking the form of a daily time series from 0 to 24 weeks, will be summarized graphically to ascertain (1) percent of intervention participants (by intervention type) that responded on any given day post baseline; (2) of those that responded, what the average response was. Polynomial smoothing plots will be used to create visualizations with basic descriptive statistics (mean, standard deviation, median, and interquartile range) used to describe text message response and workshop participation.

4.2 Analysis Plan for Aim 2

All analyses will follow an as-treated analysis method where participants are evaluated for the condition they actually received rather than the condition to which they may have originally been assigned.

Due to the effects of the COVID-19 pandemic on recruitment, intervention delivery, and potential follow-up rates, analyses will be approached from two angles. First, analyses will be performed using data from participants recruited prior to March 1, 2020, when the randomized-waitlist design was applied. The first stage of analysis will NOT include post-waitlist intervention data for those randomized to waitlist due to non-random assignment of individuals with no clear intervention preference.

Second, all data for participants recruited prior to March 1, 2020 will be pooled with those recruited after March 1, 2020 and post-waitlist data and treatment effect estimates (analogous to those performed with randomized data) will be computed with methods appropriate for observational study designs (e.g. propensity score weighting to account for treatment selection). Details of the analytic approach can be found below.

4.2.1 Analysis of trial phase data (pre-March 1, 2020 recruitment)

All analyses will adhere to CONSORT reporting guidelines[14]. All primary and secondary study outcomes are measured on continuous scales and will be reported with mean estimates and 95% confidence intervals with reference to waitlist vs non-waitlist for participants recruited prior to March 1, 2020. The estimand of interest for the primary hypotheses will be the effectiveness of MBSR, Stress Proofing, or Daily Examen for active intervention participants versus waitlist control participants, *regardless of preference*. Because trial phase data will be a mix of randomized (for those with no preference) and non-randomized (those with unique preferences) assignment to treatment, and because we expect preference for a particular intervention (or the lack of a preference for a particular intervention) to be influenced by characteristics of the individual participants which may be confounders with stress and anxiety outcomes, it is necessary to use observational/causal methods to control for such characteristics in order to generate a less biased estimate of the treatment effect.

Therefore, the main analysis will proceed in two steps:

1. Step 1. Calculation of propensity scores for assignment to each of the 3 intervention types versus the waitlist control
2. Step 2. Adjustment of the treatment effect estimation model for propensity to receive each of the intervention types.

4.2.1.1 Step 1. Calculation of propensity scores for the probability of receiving MBSR, SP, or DE (vs. assignment to the waitlist)

Calculation of propensity scores will be operationalized with the use of multinomial logistic regression with a nominal outcome of odds of being assigned to immediate receipt of MBSR, SP, DE vs. being assigned to the waitlist:[15]

$$(1)\quad ln\left(\frac{\Pr(INT_{a,i})}{\Pr(INT_{Waitlist,i})}\right) = \beta_0 + \boldsymbol{\delta} Pref_i^p + \sum_{k=1}^{K} \alpha_k X_{ik}$$

Where $INT_{a,i}$ indicates assignment of person i (i=1,….,N) to condition a (a= MBSR, SP, DE, or waitlist control); $Pref_i^p$ is a categorical variable taking 3 possible values (indifferent between 3 interventions, indifferent between 2 interventions, and unique preference for an intervention); and X_{ik} is a set of k (k=1,…K) prespecified baseline sociodemographic characteristics thought to potentially predict intervention preference, including: 1) age, sex, marital status, race, bi-vocational status, weekly work hours in the UMC, interaction between bi-vocational status and work hours, number of congregations appointed to, number of congregants served (not appointed to church, 1-149/week, 150+/week), self-endorsed anxiety, self-endorsed high cholesterol, and baseline stress/anxiety scores for C-SOSI and anxiety outcomes and 2) additionally, with caffeine intake, alcohol consumption, body mass index, physical activity[7]. Propensity scores will be calculated separately for stress/anxiety vs HRV outcomes (since HRV outcomes are available for only a subset of participants) and for trial vs. pooled participants using equation (1) with modifications based on outcome/sample as listed below:

1. Stress and anxiety outcomes: Equation (1) for **trial participants** adjusting for baseline **GAD-7 and C-SOSI scores**
2. Depression outcome: Equation (1) for **trial participants** adjusting for baseline **PHQ-8 score**

[7] An *a priori* decision given the influence that these variables exert on HRV

3. HRV outcomes: Equation (1) for **trial participants** adjusting for baseline **GAD-7, C-SOSI, PHQ-8, and HRV** outcomes
4. Stress and anxiety outcomes: Equation (1) for **pooled trial and observational participants** adjusting for baseline **GAD-7, C-SOSI scores** and trial vs. observational participants
5. Depression outcomes: Equation (1) for **pooled trial and observational participants** adjusting for baseline **PHQ-8 score** and trial vs. observational participants
6. HRV outcome: Equation (1) for **pooled trial and observational participants** adjusting for baseline **GAD7, C-SOSI, PHQ-8, and HRV** outcomes and trial vs. observational participants

Selection of sociodemographic predictors will be primarily through prior knowledge of likely correlates with intervention choice/preference/assignment but may also be queried more generally by producing summary tables of baseline characteristics (e.g. such as those included in a traditional "Table 1"), stratified by intervention assignment.

Covariate balancing propensity score methods[16] will be used to calculate propensity scores. Once initial propensity scores have been calculated, superimposed histograms will be used to visualize overlap between distributions of propensity scores for individual assigned to the 3 interventions. Balance between groups may further be ascertained by producing stratified tables that present estimates for baseline sociodemographic covariates adjusted for propensity score via regression analysis as described in Spreeuwenberg et al 2010.[15] If balance based on basic linear specification is not ideal, then covariate balancing propensity score may be employed.[16]

4.2.1.2 Step 2. Estimation of treatment effects, adjusting for propensity score for assignment to MBSR, SP, DE (vs. waitlist)

Linear mixed models will be specified with fixed effects for time and intervention status and adjustment for propensity score for each of the intervention types (vs assignment to waitlist).[17] Propensity score adjustment versus weighting is preferred because it is a validated method, less prone to issues with extreme weights, and more flexible to integration with partial cluster methods and combination with the multiple imputation process. Partially clustered random effects for the non-waitlist arms will be used to account for the clustering caused by workshop-level intervention delivery that affects participants in the immediate intervention condition. Distributions of residuals will be examined to confirm that linear modelling assumptions are met. Because timing of data collection for waitlist participants was agnostic to intervention preference, there may be potential confounding of calendar time and characteristics of individuals that inform intervention preference and/or stress/anxiety outcomes, thus calendar time must be accounted for in the analytic models. Baseline levels of C-SOSI will be visualized across calendar time of the study period to assess the presence of trends. Linear, quadratic, and cubic forms of calendar time will be tested against the continuous C-SOSI baseline level to ascertain whether non-linear trends are present with AIC used to select the best-fitting functional form, with the best fitting functional form for time included in the final regression models.

The main outcome models for Hypotheses 1-5 will be specified as follows (using cLDA structure[18]):

(2) $$Y_{ijt} = \beta_0 + \beta_1 I_{ij}^{M} * Week_{ij} + \beta_2 I_{ij}^{S} * Week_{ij} + \beta_3 I_{ij}^{D} * Week_{ij} + \beta_4 (Week_{ij} - 12)_+ + \beta_5 I_{ij}^{M} * (Week_{ij} - 12)_+ + \beta_6 I_{ij}^{S} * (Week_{ij} - 12)_+ + \beta_7 I_{ij}^{D} * (Week_{ij} - 12)_+ + \beta_8 PS_i^{M} + \beta_9 PS_i^{S} + \beta_{10} PS_i^{D} + \boldsymbol{\vartheta} f(Time) + I_{ij} u_j + \delta_i + \varepsilon_{ijt}$$

Where Y_{ijt} is a continuous outcome for reported stress or anxiety symptoms for person i in intervention delivery group j (j=1,…21) at time point t (t=0, 12, and 24 weeks post-baseline); I_{ij}^{M}, I_{ij}^{S}, and I_{ij}^{D} are indicator variables for active treatments of MBSR, Stress Proofing, and Daily Examen (respectively) for

person i in group j; $Week_{ij}$ is a continuous measure quantified as weeks from the baseline timepoint for participant i; $(Week_{ij} - 12)_+$ is a knot at 12 weeks[8], equal to $Week_{ij} - 12$ when $Week_{ij} > 12$ and 0 otherwise; PS_i^S, PS_i^M, and PS_i^D are continuous propensity scores quantifying the probability that individual i will be assigned to immediate MBSR, Stress Proofing, and Daily Examen interventions, respectively (vs. waitlist); f($Time$) is some function of calendar time in months from June 1, 2020 (either linear, quadratic or cubic based on best fit using baseline data to predict each outcome); u_j is a random intercept for the jth intervention group, and I_{ij} is an indicator that individual i in group j has been exposed to an intervention (i.e. random intercepts controlling for clustering due to group delivery are only estimated for those in active intervention conditions)[9]; δ_i is a random intercept for person i, to account for repeated measures on the same individual. We will also explore whether a random slope for time improves model fit and will include it if the difference in AIC >10 points. Robust standard errors will be used to account for the use of predicted values (propensity scores) as fixed effects in the outcome model.

Between arm comparisons will be measured by the following contrasts for Hypotheses 1-5 (using parameters specified in equation (1):

- Treatment effect (12-week) for MBSR[10]: $12 * \beta_1$
- Treatment effect (12-week) for Stress Proofing: $12 * \beta_2$
- Treatment effect (12-week) for Daily Examen: $12 * \beta_3$
- Treatment effect (24-week) for MBSR: $24 * \beta_1 + 12 * \beta_5$
- Treatment effect (24-week) for Stress Proofing: $24 * \beta_2 + 12 * \beta_6$
- Treatment effect (24-week) for Daily Examen: $24 * \beta_3 + 12 * \beta_7$

Models for HRV outcomes will use the same basic modeling structure but will omit all terms associated with the 24-week time point.

An intraclass correlation coefficient (ICC) will be reported for each outcome model, specified as follows:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\delta^2 + \sigma_\varepsilon^2}$$

Regression models for Hypothesis 6-8 will follow a similar structure as (1) but will introduce a binary indicator=1 if the patient received the intervention that they stated was their *unique* preference and 0 if the participant had no unique preference, received an intervention ranked less than first preference, or received and intervention other than the one they indicated that they uniquely preferred. Interaction terms will be introduced to allow for the calculation of separate treatment effects for those that had and received their uniquely preferred intervention vs. those without a unique preference or who did not receive their unique preference.

Specifically, parameterization will be as follows:

(3) $$Y_{ijt} = \beta_0 + \beta_1 Week_{ij} + \beta_2 Pref_i + \beta_3 I_{ij}^M * Week_{ij} + \beta_4 I_{ij}^S * Week_{ij} + \beta_5 I_{ij}^D * Week_{ij} + \beta_6 Pref_i * I_{ij}^M * Week_{ij} + \beta_7 Pref_i * I_{ij}^S * Week_{ij} + \beta_8 Pref_i * I_{ij}^D * Week_{ij} + \beta_9 (Week_{ij} - 12)_+ + \beta_{10} I_{ij}^M * (Week_{ij} - 12)_+ + \beta_{11} I_{ij}^S * (Week_{ij} - 12)_+ + \beta_{12} I_{ij}^D * (Week_{ij} - 12)_+ + \beta_{13} Pref_i * I_{ij}^M * (Week_{ij} - 12)_+ + \beta_{14} Pref_i * I_{ij}^S * (Week_{ij} - 12)_+ + \beta_{15} Pref_i * I_{ij}^D * (Week_{ij} - 12)_+ + \beta_{16} PS_i^M + \beta_{17} PS_i^S + \beta_{18} PS_i^D + \boldsymbol{\vartheta} f(Time) + I_{ij} u_j + \delta_i + \varepsilon_{ijt}$$

[8] Note that an a priori decision was made to model time from baseline as a continuous variable with a linear spline based on knowledge that significant variation in timing of surveys occurred, particularly for a subset of individuals who may have entered the intervention period before their waitlist time was complete.

[9] See "Additional File" from (*Flight L, Allison A, Dimairo M, Lee E, Mandefield L, Walters SJ. Recommendations for the analysis of individually randomised controlled trials with clustering in one arm - a case of continuous outcomes. BMC Med Res Methodol. 2016;16(1):165. doi:10.1186/s12874-016-0249-5)* for suggestions on coding of partial clustering term.

[10] An alternative, potentially unbiased measure of the treatment effect would be to measure treatment effects only for participants that did not have a unique preference for a particular intervention[19,20]

Where $Pref_i$ is an indicator =1 if individual i enrolled in the intervention type that was their clear preference (ranked first, with no ties) and 0 otherwise.

For the waitlist participants in the pooled trial and observational analysis, it will be possible for the preference variable to take on a 1 value during the waitlist and a 0 value during post-waitlist in the special case where an individual initially expressed a unique preference, was randomized to the waitlist, and then received an intervention other than their initially indicated preferred intervention.

Between arm comparisons will be measured by the following contrasts for Hypothesis 6-8 (using parameters specified in equation (2):

- For MBSR:
  - Treatment effect (12-week) for those with no #1 preference[11]: $12 * \beta_3$
  - Treatment effect (12-week) for those that received their #1 preference: $12 * \beta_3 + 12 * \beta_6$
  - Difference in treatment effect between preference/no preference (12-week): $12 * \beta_6$
  - Treatment effect (24-week) for those with no #1 preference: $24 * \beta_3 + 12 * \beta_{10}$
  - Treatment effect (24-week) for those that received their #1 preference: $24 * \beta_3 + 24 * \beta_6 + 12 * \beta_{10} + 12 * \beta_{13}$
  - Difference in treatment effect between preference/no preference (24-week): $24 * \beta_6 + 12 * \beta_{13}$
- For Stress Proofing:
  - Treatment effect (12-week) for those with no #1 preference[8]: $12 * \beta_4$
  - Treatment effect (12-week) for those that received their #1 preference: $12 * \beta_4 + 12 * \beta_7$
  - Difference in treatment effect between preference/no preference (12-week): $12 * \beta_7$
  - Treatment effect (24-week) for those with no #1 preference: $24 * \beta_4 + 12 * \beta_{11}$
  - Treatment effect (24-week) for those that received their #1 preference: $24 * \beta_4 + 24 * \beta_7 + 12 * \beta_{11} + 12 * \beta_{14}$
  - Difference in treatment effect between preference/no preference (24-week): $24 * \beta_7 + 12 * \beta_{14}$
  -
- For Daily Examen:
  - Treatment effect (12-week) for those with no #1 preference[8]: $12 * \beta_5$
  - Treatment effect (12-week) for those that received their #1 preference: $12 * \beta_5 + 12 * \beta_8$
  - Difference in treatment effect between preference/no preference (12-week): $12 * \beta_8$
  - Treatment effect (24-week) for those with no #1 preference: $24 * \beta_5 + 12 * \beta_{12}$
  - Treatment effect (24-week) for those that received their #1 preference: $24 * \beta_5 + 24 * \beta_8 + 12 * \beta_{12} + 12 * \beta_{15}$
  - Difference in treatment effect between preference/no preference (24-week): $24 * \beta_8 + 12 * \beta_{15}$

An average effect of having a preference can be computed by taking the average (by linear combination) of the treatment effects across all 3 interventions for those with a clear preference, then the average treatment effect for those without a clear preference, and the difference between those averages. This is referred to as a "selection effect" (as detailed in previous literature[7]) and should not be confused with preference effects, which measures the effects of receiving a preferred intervention versus receiving an intervention which is not preferred (we cannot measure the latter quantity given the partially randomized preference design). Selection effects measure the difference in outcomes between participants who would select one treatment or the other if they were free to do so and those that would not.

[11] Note: This measure has been proposed as an alternative and potentially unbiased measurement for the treatment effect though measurement will be with less precision because of smaller sample size.[19,20]

For the primary outcomes of stress symptoms and HRV MESOR, hypothesis tests will be considered individual testing for each of the 3 intervention types and multiple outcomes to be considered conjunction testing. Therefore p-values will be calculated for between arm differences at 12-weeks adjusting for two hypotheses for each of the 3 interventions[21]. Adjustments will use the Benjamini-Hochberg correction[22] (raw p-values may also be made available), which is more appropriate than conservative approaches such as the Bonferroni method when outcomes are correlated. Adjusted p-values will only be presented for primary outcomes.

4.2.2 Sensitivity Analysis: Analysis of pooled observational and randomized trial data (pre- and post-March 1, 2020 recruitment)

4.2.2.1 *Step 1. Calculation of propensity scores for the probability of receiving MBSR, SP, or DE (vs. waitlist)*

Calculation of propensity scores for receipt of interventions (vs. not having a unique preference) will proceed in the same manner when trial and observational data are pooled as when only the pre-March 1, 2020 cohort (without post-waitlist) data was used, with the addition of an indicator variable for post-March 1, 2020 study entry (anecdotally, post-March 1, 2020 entrants may have been influenced by trial participants to participate in certain interventions). In addition, participants randomized to the waitlist (who subsequently enter an intervention phase) will have two propensity scores calculated (via having two observations in the baseline dataset used to calculate propensity scores), one for the baseline observation from their waitlist period and one for the baseline observation of their intervention period. Regression specification for step 1 will otherwise will mirror that employed for the trial-phase cohort.

4.2.2.2 *Step 2. Analysis models adjusting for propensity scores*

Analyses pooling all data will share virtually the sample regression specification with the addition of an indicator for probability of being in the observational (post-March 1, 2020 cohort or post-waitlist).

The main outcome models for Hypotheses 1-5 will be specified with the same model specification, with a binary indicator (Obs_i) for post-March 1, 2020 entrance into the study:

(1) $$Y_{ijt} = \beta_0 + \beta_1 I_{ij}^{M} + \beta_2 I_{ij}^{S} + \beta_3 I_{ij}^{D} + \beta_4 Week_{ij} + \beta_5 I_{ij}^{M} * Week_{ij} + \beta_6 I_{ij}^{S} * Week_{ij} + \beta_7 I_{ij}^{D} * Week_{ij} + \beta_8 (Week_{ij} - 12)_{+} + \beta_9 I_{ij}^{M} * (Week_{ij} - 12)_{+} + \beta_{10} I_{ij}^{S} * (Week_{ij} - 12)_{+} + \beta_{11} I_{ij}^{D} * (Week_{ij} - 12)_{+} + \beta_{12} PS_i^{M} + \beta_{13} PS_i^{S} + \beta_{13} PS_i^{D} + \boldsymbol{\vartheta} f(Time) + \beta_{14} Obs_i + I_{ij} u_j + \delta_i + \varepsilon_{ijt}$$

Binary indicators for intervention received will denote that a participant is in the active exposure period. The covariate for within study time (i.e. time from baseline rather than months from June 1, 2020) will indicate weeks from baseline. Importantly, study participants that were part of the waitlist may have time points (in weeks from baseline) that may be measured multiple times due to having both a waitlist and an active condition, for illustration, the following are example data from a waitlist trial participant and an observational participant:

Participant	Study time (weeks)	In intervention period?	Calendar time (months from June 1, 2020)	Cohort
1	0	No	1	Trial
1	12	No	3	Trial

1	24	No	6	Trial
1	0	Yes	8	Trial
1	12	Yes	11	Trial
1	24	Yes	14	Trial
2	0	Yes	17	Observational
2	12	Yes	20	Observational
2	24	Yes	23	Observational

We may perform sensitivity analyses adjusting for any influential covariates (e.g. baseline financial stress, depressive symptoms, marital status; see Table 2) that are imbalanced by chance between waitlist and non-waitlist arms.

4.2.3 Missing data

Because randomization occurred prior to baseline data collection, it is possible for participants to be randomized and then not provide any study data, including baseline data. This may lead to selection bias and imbalance in participant characteristics that may correlate with treatment preference/assignment, and may also influence the outcome. Our complete-case analytic sample entry criteria will be that participants provided at least the baseline assessment, with "missing data" defined as any subsequent missing survey outcome (or covariate) data at 12- or 24-weeks. However, we will consider weighting our analytic sample by the probability of selection, aided by registration data and/or CHI Panel 2019 data if there is sufficient overlap CHI Panel 2019 data with missing baseline data.

If 5% or more of participants miss follow-up data, a sensitivity analysis will be performed using multiply imputed data (using chained equations, aka MICE[23,24]) to account for potentially informative missingness and estimates will be compared to the original complete case estimates. This will be performed in a 4-step manner:

1. Imputation using MICE for all variable included in either the final regression model OR the propensity score model
2. Export of the imputed dataset containing variables used for propensity score calculation followed by separate propensity score calculations for each MI dataset
3. Re-import of propensity scores for each intervention and MI dataset back to the original dataset (followed by flagging of propensity scores as super-varying variables)
4. Final regression analyses run and combined using Rubin's rules

Note that analysis with imputed datasets will be considered a sensitivity analysis, focused on ascertaining whether there was bias caused by informative missingness in study dropout patterns, thus focus will be more on the magnitude of estimated effect than the variance. Further, due to software limitations and small sample size, it will not be possible to account for partial clustering (intervention group) and repeated measures (multiple time periods for the same person) in the MI process, thus we may expect standard errors to be *underestimated*.

Risk Management Strategies

Adverse outcomes will be monitored through participant and instructor reporting to the research team and through in-depth interviews.

5 Limitations

Though partially randomized preference trials may enhance external validity of a study,[25] estimates of treatment effects may suffer from similar biases as may be seen in an observational study due to confounding between characteristics that give rise to preference and the outcome under study. Some previous studies suggest that

analyzing study outcomes for the subsample of participants who were indifferent to their intervention allocation can provide an unbiased treatment effect estimate,[20] though precision of measurement is lost due to smaller sample size. In the case of this study, the randomization occurred prior to baseline data collection and any study dropout before baseline may lessen the effectiveness of the randomization for providing unbiased data for calculation of treatment effects. Another approach is to control for confounding characteristics between preference and the outcome which may lead to estimates with more precision and an relatively unbiased estimate of the treatment effect.[19] In this study, we hope to use both approaches in order to reinforce confidence in the accuracy of treatment effect estimates. However, the propensity score adjustment approach will only correct confounding bias if the propensity model is well specified[17] which is inherently difficult to definitively prove. Finally, due to randomization prior to baseline data collection, we can not rule out the possibility that knowledge of treatment assignment affected baseline outcome measurements..

6 Appendix

6.1 Supporting Tables and Charts

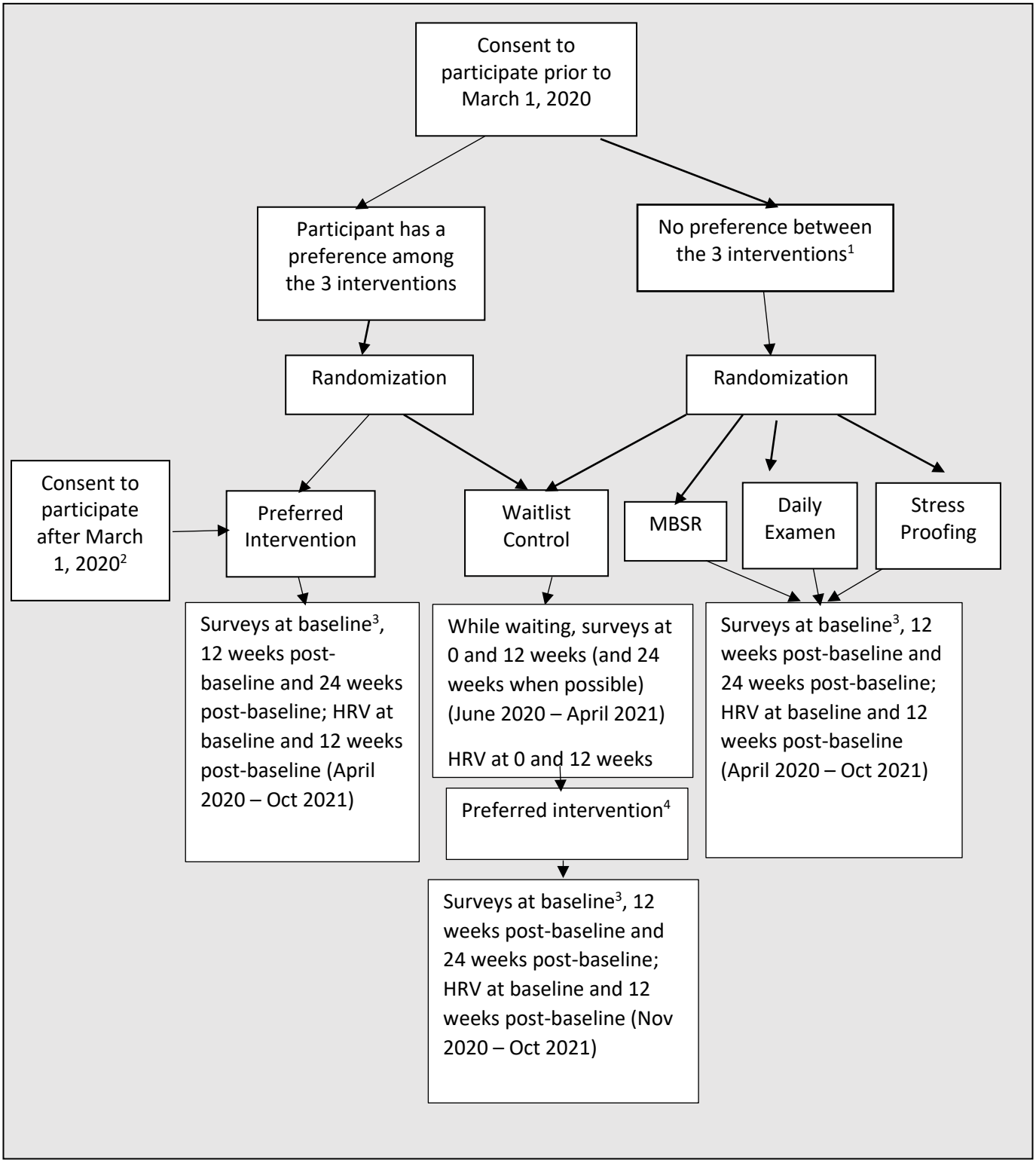
6.1.1 Figure 1. Pandemic-adapted Selah Study Design: A Partially-randomized Waitlist-controlled Preference Trial

[1] Participants who prefer 2 interventions equally and over the third will be first randomly assigned between the 2 preferred interventions, and then randomized into either the non-waitlist arm with their preferred intervention or the waitlist arm.

[2] All participants who consented after March 1, 2020 were non-randomly assigned to their preferred intervention. Participants with no preference selected a workshop with dates of their choosing.

[3] Baseline indicates immediately before intervention start.

[4] After waitlist participants finish giving control data, they may participate in the intervention they originally indicated as their preference or change it to reflect their current preference.



[1] Participants who prefer 2 interventions equally and over the third will be first randomly assigned between the 2 preferred interventions, and then randomized into either the non-waitlist arm with their preferred intervention or the waitlist arm.

[2] All participants who consented after March 1, 2020 were non-randomly assigned to their preferred intervention. Participants with no preference selected a workshop with dates of their choosing.

[3] Baseline indicates immediately before intervention start.

[4] After waitlist participants finish giving control data, they may participate in the intervention they originally indicated as their preference or change it to reflect their current preference.

7 References

1. Baruth M, Wilcox S, Evans R. The health and health behaviors of a sample of African American pastors. *J Health Care Poor Underserved*. 2014;25(1):229-241. doi:10.1353/hpu.2014.0041

2. Knox S, Virginia SG, Lombardo JP. Depression and Anxiety in Roman Catholic Secular Clergy. *Pastoral Psychology*. May 2002.

3. Ferguson TW, Andercheck B, Tom JC, Martinez BC, Stroope S. Occupational conditions, self-care, and obesity among clergy in the United States. *Soc Sci Res*. 2015;49:249-263. doi:10.1016/j.ssresearch.2014.08.014

4. Varvogli L, Darviri C. Stress Management Techniques: evidence-based procedures that reduce stress and promote health. *Health Science Journal*. 2011;5(2).

5. Rao NP, Varambally S, Gangadhar BN. Yoga school of thought and psychiatry: Therapeutic potential. *Indian J Psychiatry*. 2013;55(Suppl 2):S145-9. doi:10.4103/0019-5545.105510

6. Spring B. Evidence-based practice in clinical psychology: what it is, why it matters; what you need to know. *J Clin Psychol*. 2007;63(7):611-631. doi:10.1002/jclp.20373

7. Walter SD, Turner R, Macaskill P, McCaffery KJ, Irwig L. Beyond the treatment effect: Evaluating the effects of patient preferences in randomised trials. *Stat Methods Med Res*. 2017;26(1):489-507. doi:10.1177/0962280214550516

8. Sidani S, Epstein DR, Bootzin RR, Moritz P, Miranda J. Assessment of preferences for treatment: validation of a measure. *Res Nurs Health*. 2009;32(4):419-431. doi:10.1002/nur.20329

9. Kabat-Zinn J, Hanh TN. *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. Delta; 2009.

10. Case AD, Keyes CLM, Huffman KF, et al. Attitudes and behaviors that differentiate clergy with positive mental health from those with burnout. *J Prev Interv Community*. 2020;48(1):94-112. doi:10.1080/10852352.2019.1617525

11. Meichenbaum D. *Stress inoculation training: A preventative and treatment approach.* (In P. M. Lehrer RLW, ed.). New York, NY: Guilford Press; 2007.

12. Meichenbaum D, Cameron R. Stress Inoculation Training. In: Meichenbaum D, Jaremko ME, eds. *Stress reduction and prevention*. Boston, MA: Springer US; 1989:115-154. doi:10.1007/978-1-4899-0408-9_5

13. Carlson LE, Thomas BC. Development of the Calgary Symptoms of Stress Inventory (C-SOSI). *Int J Behav Med*. 2007;14(4):249-256. doi:10.1007/BF03003000

14. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18. doi:10.1186/1741-7015-8-18

15. Spreeuwenberg MD, Bartak A, Croon MA, et al. The Multiple Propensity Score as Control for Bias in the Comparison of More Than Two Treatment Arms: An Introduction From a Case Study in Mental Health on JSTOR. *Med Care*. 2010;48(2):166-174.

16. Imai K, Ratkovic M. Covariate balancing propensity score. *J Royal Statistical Soc B*. 2014;76(1):243-263. doi:10.1111/rssb.12027

17. Vansteelandt S, Daniel RM. On regression adjustment for the propensity score. *Stat Med*. 2014;33(23):4053-4072. doi:10.1002/sim.6207

18. Coffman CJ, Edelman D, Woolson RF. To condition or not condition? Analysing "change" in longitudinal randomised controlled trials. *BMJ Open*. 2016;6(12):e013096. doi:10.1136/bmjopen-2016-013096

19. Gemmell I, Dunn G. The statistical pitfalls of the partially randomized preference design in non-blinded trials of psychological interventions. *Int J Methods Psychiatr Res*. 2011;20(1):1-9. doi:10.1002/mpr.326

20. Walter SD, Bian M. Relative efficiencies of alternative preference-based designs for randomised trials. *Stat Methods Med Res*. 2020;29(12):3783-3803. doi:10.1177/0962280220941874

21. Rubin M. When to adjust alpha during multiple testing: a consideration of disjunction, conjunction, and individual testing. *Synthese*. July 2021. doi:10.1007/s11229-021-03276-4

22. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Statist*. 2001;29(4):1165-1188. doi:10.1214/aos/1013699998

23. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol*. 2010;171(5):624-632. doi:10.1093/aje/kwp425

24. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007;16(3):219-242. doi:10.1177/0962280206074463

25. TenHave TR, Coyne J, Salzer M, Katz I. Research to improve the quality of care for depression: alternatives to the simple randomized clinical trial. *Gen Hosp Psychiatry*. 2003;25(2):115-123. doi:10.1016/S0163-8343(02)00275-X