A Two-Year Holistic Health and Stress Intervention

Results of a Randomized Controlled Trial in Clergy

Rae Jean Proeschold-Bell, PhD
Duke Global Health Institute, and Duke Center for Health Policy & Inequalities Research
Duke University, Durham, NC

Elizabeth L. Turner, PhD
Department of Biostatistics & Bioinformatics, and Duke Global Health Institute
Duke University, Durham, NC

Gary G. Bennett, PhD
Department of Psychology & Neuroscience, and Duke Global Health Institute
Duke University, Durham, NC

Jia Yao, MA
Duke Global Health Institute, and Center for Health Policy & Inequalities Research
Duke University, Durham, NC

Xiang-Fang Li, MD, MPH
Duke Divinity School, Durham, NC

David E. Eagle, PhD
Duke Global Health Institute, and Center for Health Policy & Inequalities Research
Duke University, Durham, NC

Rachel A. Meyer, BA
Duke Divinity School, Durham, NC

Redford B. Williams, MD
Departments of Psychiatry & Behavioral Sciences, and Psychology & Neuroscience, Medicine, and Behavioral Medicine Research Center
Duke University, Durham, NC

Abstract

This study sought to determine the effect of a two-year, multi-component health intervention (Spirited Life) targeting metabolic syndrome and stress simultaneously.

Design A randomized controlled trial using a three-cohort multiple-baseline design was conducted in 2010-2014.

Setting/Participants Participants were United Methodist clergy in North Carolina, USA, in 2010, invited based on occupational status. Of 1,745 clergy invited, 1,114 consented, provided baseline data and were randomly assigned to immediate-intervention (n=395), one-year waitlist (n=283), or two-year waitlist (n=436) cohorts for a 48-month trial duration.

Intervention The two-year intervention consisted of personal goal-setting and encouragement to engage in monthly health coaching, an online weight-loss intervention, a small grant, and three workshops delivering stress management and theological content supporting healthy behaviors. Participants were not blinded to intervention.

Main Outcome Measures Trial outcomes were metabolic syndrome (primary) and self-reported stress and depressive symptoms (secondary). Intervention effects were estimated in 2016 in an intention-to-treat framework using generalized estimating equations (GEE) with adjustment for baseline level of the outcome and follow-up time points. Log-link Poisson GEE with robust standard errors was used to estimate prevalence ratios (PR) for binary outcomes; mean differences were used for continuous/score outcomes.

Results Baseline prevalence of metabolic syndrome was 50.9% and depression was 11.4%. The 12-month intervention effect showed a benefit for metabolic syndrome (PR: 0.86; 95% confidence interval [CI]: 0.79, 0.94; $p<0.001$). This benefit was sustained at 24 months of

Robin Y. Swift, MPH
Duke Divinity School, Durham, NC

H. Edgar Moore, PhD
Duke Divinity School, Durham, NC

Melanie A. Kolkin, BA
Duke Divinity School, Durham, NC

Carl C. Weisner, MA
Duke Divinity School, Durham, NC

Katherine M. Rugani, MS
Duke Divinity School, Durham, NC

Holly J. Hough, PhD
Duke Divinity School, Durham, NC

Virginia P. Williams, PhD
Williams LifeSkills, Inc., Durham, NC

David C. Toole, PhD, MTS, MPH
Duke Divinity School and Duke Global Health Institute
Duke University, Durham, NC

Please direct comments to Rae Jean Proeschold-Bell at rae.jean@duke.edu.

intervention (PR: 0.88; 95% CI: 0.78, 1.00; p=0.04). There was no significant effect on depression or stress scores.

Conclusions The Spirited Life intervention improved the metabolic syndrome prevalence in a population of U.S. Christian clergy and sustained improvements during 24 months of intervention. These findings offer support for long-duration behavior change interventions and population-level interventions that allow participants to set their own health goals.

The is the peer-reviewed version of the following article: Proeschold-Bell, R.J., Turner, E. L., Bennett, G. G., Yao, J., Li, X.-F., Eagle, D. E., Meyer, R. A., Williams, R. B., Swift, R. Y., Moore, H. E., Kolkin, M. A., Weisner, C. C., Rugani, K. M., Hough, H. J., Williams, V. P., & Toole, D. C. (2017, June). A 2-year holistic health and stress intervention: Results of an RCT in clergy. American Journal of Preventive Medicine, 53(3), 290-299. doi: 10.1016/j.ampere.2017.04.009, which is published in final form at: http://www.ajpmonline.org/article/S0749-3797(17)30244-1/fulltext

Introduction

The metabolic syndrome (MetS), defined by the International Diabetes Federation as central obesity plus any two of the following: elevated triglycerides, low high-density lipoprotein, hypertension, and abnormal glucose regulation,[1] is associated with increased risk for type 2 diabetes, cardiovascular disease, stroke, and mortality.[1,2,3] Weight loss is one approach to reverse MetS. Although weight loss trials have demonstrated decreases in weight, improvements are rarely maintained at 12 and 24 months post-intervention.[4,5] One hypothesis as to why weight improvements are temporary is that chronic stress might drive, via elevation in glucocorticoid secretion, a desire to consume caloric, energy-dense food. Consumption of comfort foods may stimulate pleasure centers in the brain, thus regulating stress-induced systemic arousal.[6] Longer interventions may allow participants to practice dietary behaviors during both high- and low-stress periods. Such long-term practice under diverse conditions may be critically important to sustaining weight loss.

Potentially, jointly targeting MetS and stress management for two years may reduce MetS and decrease weight due to stress management and long-term practice of healthy dietary behaviors. Few studies have examined interventions with dual primary aims of weight loss and stress management, and they have been relatively small, short-term pilot studies with highly selective, mostly female samples.[7,8] Outcomes generally have been positive, but due to short follow-up periods, have not addressed the challenge of long-term behavior change.

One population that suffers chronic stress and high rates of obesity is clergy. Clergy experience a number of work-related stressors, including work overload, unpredictable schedules, intrusiveness, and criticism from parishioners.[9] Clergy exhibit above-average rates of depression and obesity.[10,11] Obesity prevalence in United Methodist Church (UMC) clergy was 41% in a national U.S. study and also in a North Carolina (NC) study.[12,13]

There are several benefits to studying obesity, stress, and MetS in clergy. First, clergy have high rates of obesity and chronic stress. Second, a large percentage of clergy are male; among UMC clergy, approximately 71% are male.[14] In spite of the fact that male obesity rates appear to be climbing,[15] men have been under-represented in weight loss interventions to date.[16] Third, successful interventions tailored to Christian clergy, estimated at 244,200 in the U.S.,[17] may be adapted and offered to the large number of Christian churchgoers in the U.S.

The Spirited Life trial was a pragmatic trial (estimated PRECIS score of 84% with 100% being extremely pragmatic;[18,19] see Supplement) of a combined weight reduction and stress management intervention among an employee population of clergy. It was designed to assess changes in the prevalence of MetS (primary outcome), weight, depression, and stress symptoms (secondary outcomes). Details of the trial rationale, intervention, and implementation are available elsewhere.[20] The trial used a multiple baseline trial design with three cohorts randomly assigned to intervention start dates spaced one year apart (immediate-intervention, one-year waitlist and two-year waitlist cohorts). The primary hypothesis was that the intervention would lead to a lower prevalence (or mean level) of MetS, weight, stress symptoms, and depression at 12, 18, and 24 months of intervention. The 12-, 18- and 24-month intervention effects were estimated using standard modeling approaches for data from a multiple baseline RCT. This article reports outcomes during the three cohorts' intervention and waiting periods.

Methods

The original protocol for the trial has been published.[20] The CONSORT checklist is provided as supporting information (Tables S2-S3).

Study population

Eligible participants were all clergy members in July 2010 of the NC Annual Conference and the Western NC Annual Conference of the United Methodist Church; these two governing bodies employ approximately 1,800 UMC clergy. All individuals were invited based on clergy occupation status rather than health status. There were no health status inclusion criteria, and clergy with and without MetS, depression, and stress symptoms were recruited. Exclusion criteria were intentionally few; pastors on leave and most extension ministers (e.g., seminary professors, hospital chaplains) were excluded.

Procedure

An extensive communication campaign was conducted in September-October 2010 to inform all NC UMC clergy, regardless of health status, about the trial. 1,745 eligible clergy were invited to participate, beginning with online consent. Consenting participants had to complete both an in-person cardiometabolic screening and online survey to enroll. A total of 1,114 clergy (64%) met these criteria. Using a randomized multiple baseline design, participants were randomly assigned to one of three intervention start dates: January 2011 (Cohort 1, "immediate-intervention cohort"); January 2012 (Cohort 2, "one-year waitlist cohort"); and January 2013 (Cohort 3, "two-year waitlist cohort"). Start dates were spaced one year apart so that all three cohorts began the two-year intervention during the same season. More participants (40%) were randomized to the two-year waitlist cohort to guard against attrition. More participants were randomized to the immediate-intervention cohort (35%) than the one-year waitlist cohort (25%) to intervene with more clergy sooner. Randomization was stratified by geographic district (sub-administrative units in Conferences: 27 levels). Using the list of 1,114 enrolled participants, an independent statistician generated random allocation sequences for each district, ensuring the overall 40%:35%:25% split. Random allocations were implemented by study personnel. Due to the pragmatic nature of the trial, blinding of participants and intervention personnel was not possible. Personnel who measured physical outcomes were blinded to trial cohort.

The two-year Spirited Life intervention began with a required three-day workshop. Figure 1 shows the flow of participants through randomization and data collection. Cardiometabolic and survey data were collected for the entire sample at the trial baseline (fall 2010) and repeatedly through fall 2014. Each cohort was assessed just prior to intervention and at 12, 18, and 24 months into the intervention. In addition, the two-year waitlist cohort was assessed at every time point that the immediate-intervention cohort was assessed to provide control comparison measurements for the full two-year intervention duration of the immediate-intervention cohort (Figure 1). Duke University's Arts & Sciences Institutional Review Board (IRB) approved all procedures and participants gave free and informed consent.

Intervention

The intervention consisted of four components, described here briefly and elsewhere in detail.[20] Only the initial workshop was required. It delivered the Williams LifeSkills (WLS) stress management program plus theological content supporting healthy behaviors (e.g., God's becoming flesh in Jesus urges Christians to be good stewards of their bodies). WLS is a protocol-driven, manualized training program shown to improve stress

coping and interpersonal relationship skills.[21] Two additional two-day workshops were spaced mid-way and at the end of the intervention. They included opportunities for clergy to articulate core values, re-commit to behavior change, and plan for sustaining their accomplishments. Intervention health coaches contacted participants after their initial workshop to schedule health coaching calls. Participants were encouraged to have monthly calls but were allowed to space them less frequently. During calls, health coaches utilized motivational interviewing with a focus on goal-setting and support (see Supplement). Regardless of weight status, participants were encouraged to register for a ten-week, online, weight-loss program called Naturally Slim®. Naturally Slim® emphasized eating only when hungry; decreasing sugar intake; eating smaller portions; and balancing fats, proteins, and carbohydrates. In January of the second year, participants were encouraged to apply for $500 grants to assist in achieving their health goals. The intervention content was the same for each cohort, except the two-year waitlist cohort was offered the online stress management program meQuilibrium (https://www.mequilibrium.com/) rather than WLS, based on participant feedback that clergy training includes much of the WLS content. Both WLS and meQuilibrium have cognitive behavioral underpinnings. WLS focused on deciding between action and deflection, problem-solving, assertion, listening, and empathy with many role plays, whereas meQuilibrium offered self-assessments of one's environment, interpersonal relationships, and the thoughts that precede emotions, paired with online journaling and exercises.

Measures

Cardiometabolic data collection, performed by staff trained using detailed protocols (see Supplement), assessed the five MetS components. MetS indicators were derived for each participant at each measurement time point using the International Diabetes Federation definition1 (see Supplement). Body Mass Index (BMI) categories were created using the National Heart, Lung, and Blood Institute definition.[22]

A 45-minute online survey included the secondary outcome measures of stress symptoms and depression. Stress symptoms were measured using the 10-item Perceived Stress Scale (PSS), with scores ranging from 0 to 40. Depressive symptoms were measured using the Patient Health Questionnaire-8 (PHQ-8),[23] consisting of eight items on the frequency of depression symptoms during the past two weeks, with scores ranging from 0 to 24. Based on previous validation studies, scores of 10 or higher were used to indicate moderate or severe depression[24] (referred to as "depression").

Outcomes

The pre-specified primary outcome measure was prevalence of MetS and secondary outcome measures were the prevalence of depression (PHQ-8 ≥10), mean stress scores, and mean weight, comparing the immediate-intervention and the two-year waitlist cohort 24 months after trial baseline. Additional pre-specified comparisons included the same comparisons at 12 months after trial baseline (see Supplement). Data from all three cohorts (waitlist and intervention periods) were combined in a single statistical model. Combining all information available in the intervention periods of both waitlist cohorts maximized statistical power to estimate the 12-, 18- and 24-month intervention effect for each outcome (see Data Analyses).

Statistical power

The trial was powered at 83% to detect a difference of 10 percentage points for MetS prevalence, and powered at 78% for a 5.6 percentage-point difference for depression prevalence, between immediate-

intervention and two-year waitlist cohorts at 24 months using a two-tailed t-test at the 5% significance level. See Supplement for additional power analyses.

Statistical analysis

Baseline and follow-up data were summarized by randomized cohort as appropriate: cases (proportions) for categorical outcomes and means (standard deviations [SD]) for continuous outcomes. The intention-to-treat principle was used for all follow-up analyses, whereby all participants were analyzed in the cohort to which they were randomized even if they later changed cohort or did not participate in intervention activities at any time. All reported p-values are two-sided. Analyses were based on a pre-specified analysis plan and performed using SAS (version 9.4) and Stata (version 14.1) in 2016.

Data from the three cohorts at all follow-up time points and from all participants were modeled together, including data from participants who were later lost to follow-up. To account for within-person correlation of outcomes due to multiple follow-up measures on each participant, generalized estimating equations (GEE) were used to estimate population-averaged effect estimates.[25] An unstructured correlation matrix with robust standard errors was used to account for the correlation between multiple responses for the same participants. To estimate prevalence ratios (PR) for binary outcomes (MetS, MetS components, depression, and attained target proportion of weight loss), a Poisson distribution with log-link – a valid approach for binary outcomes when used in the GEE framework with robust standard errors – was used.[26] A Gaussian regression with an identity link for continuous outcomes (weight) and score outcomes (perceived stress) was used to estimate mean differences. Robust standard errors were used to account for possible model misspecification (e.g. due to slight skewness). All models treated intervention level (4 levels: waiting, 12-months, 18-months and 24-months of intervention) and post-baseline follow-up time point (7 levels: 6-month intervals from 12 months to 48 months post-baseline – see Figure 2) as categorical factors. Post-baseline follow-up time point was included to account for the possible confounding effect of time that is due to naturally occurring health changes across such a lengthy (48-month) trial.[27] All models adjusted for the baseline level of the outcome and district as a categorical factor (27 levels) to account for the stratified randomization.[28]

Stratified analyses were conducted to examine weight separately for participants who were obese and overweight at baseline. To assess clinical benefit, individual weight loss of 3% and 5% of the starting weight for all participants was examined, as even modest weight loss may improve blood pressure, cholesterol and blood sugar levels.[29,30,31,32] GEE analysis of all available data provides unbiased estimated intervention effects when the outcome missing data pattern is either missing completely at random or a covariate-dependent missing pattern,[33] and the predictors of missing outcomes are included as covariates. Doubly robust multiple imputation was performed to test whether the results were robust to missing data.[34] The use of this procedure did not substantively alter the results (see Supplement Table 1).

Results

Sample and follow-up characteristics (Table 1, Figure 1)

Participants were predominantly male (69.3%), white (89.0%), married (89.0%) and obese or overweight (82.8%) with a mean (SD) age of 51.9 (10.0) years (Table 1). The three cohorts were comparable at baseline for MetS, with an overall prevalence of 50.9%. A higher proportion (15.1%) of participants in the immediate-intervention cohort were classified as having depression compared to the one-year (11.0%) and two-year waitlist (8.4%) cohorts. Figure 1 shows that 26 participants withdrew or died before the first follow-up (12 months). Overall, 1,054 (94.6%) participants provided at least one follow-up measurement. Baseline outcomes indicated that,

compared to the 1,054 who provided at least one follow-up measurement, the 60 (5.4%) participants with no follow-up data were more likely at baseline to have depression (15.0% vs. 11.2%), MetS (60.0% vs. 50.3%), and hypertension (66.7% vs. 51.8%), but not central obesity, elevated triglycerides, elevated HbA1c, or reduced HDL. The 60 participants with no follow-up data were spread across cohorts (n=28, 24, and 8 for the three respective cohorts). Sensitivity analyses including the lost cases with imputed values for MetS indicated the results were robust to missing values (see Supplement).

For all cohorts, response rates, shown in Figure 1, exceeded 75% in the first 24 months of the trial. The lowest response rate was 69% for the two-year waitlist cohort's 48-month cardiometabolic measurement.

MetS Outcomes (Tables 1, 2 and S4; Figure 2)

Baseline prevalence of MetS was 50.9% for the whole trial sample (Table 1). Changes in observed and model-based estimates of MetS prevalence by cohort over time are shown in Supplement Table 3 and Figure 2, respectively. For those with at least one follow-up measurement, there were decreases in MetS prevalence in each cohort (ranging from 3.7 to 6.6 percentage points) from immediately pre-intervention to 24 months of intervention (Table S4). These changes were 49.5% to 42.9% for immediate-intervention; 49.8% to 46.1% for one-year waitlist; and 49.6% to 45.1% for two-year waitlist cohorts. Using all intervention and control period data from all cohorts, and adjusting for follow-up time points, the 12-month intervention effect on the primary outcome of MetS (Table 2) was estimated to have 14% lower prevalence (PR: 0.86; 95% confidence interval [CI]: 0.79, 0.94, p<0.001). This effect was sustained over two years with a 24-month intervention effect estimated at a lower prevalence of 12% (PR: 0.88; 95% CI: 0.78, 1.00, p=0.04).

Components of MetS Outcomes (Tables 1, 2 and S4)

Baseline prevalence of the five MetS components in the whole sample are shown in Table 1. The most prevalent components at baseline were central obesity (81.2%), low HDL (57.4%) and hypertension (52.6%), and the less prevalent components were elevated triglycerides (50.9%) and abnormal glucose regulation (13.7%). Prevalence of the five components by cohort over time are shown in Table S4. Using all intervention and control period data from all cohorts and adjusting for follow-up time points, there was a beneficial 24-month intervention effect for the three most prevalent components with PR for central obesity of 0.91 (95% CI: 0.86, 0.96; p<0.001), for low HDL of 0.90 (95% CI: 0.81, 1.00; p=0.04), and for hypertension of 0.81 (95% CI: 0.72, 0.91; p<0.001) (Table 2). Comparable benefits were estimated at 12 months for all three outcomes, together with a benefit for elevated triglycerides (PR: 0.87; 95% CI: 0.79, 0.96; p=0.005), which was not sustained at 24 months of intervention (PR: 0.96; 95% CI: 0.84, 1.09; p=0.53).

Weight Outcomes (Tables S5-S7)

Beneficial weight outcomes were found for each cohort and time point. From immediately pre-intervention to 24 months of intervention, weight change was -3.4kg for immediate-intervention, -4.4kg for one-year waitlist, and -1.7kg for two-year waitlist cohorts (Table S5). The overall 24-month intervention effect was estimated as a mean weight of 1.75kg (95% CI: 0.74, 2.76; p<0.001) less than control, and, for participants who were obese at baseline, was 1.81kg (95% CI: 0.01, 3.62; p=0.048) less (Table S7). The Supplement and Table S6 report more weight outcomes, including loss of three and five percent of baseline body weight.

Depression and Stress Outcomes (Tables 1, 2 and S4)

Baseline prevalence of depression was 11.4% across the whole trial sample (Table 1). Changes in prevalence of depression by cohort over time are shown in Table S4. There was no evidence of an intervention benefit on depression with a 12-month PR of 1.03 (95% CI: 0.78, 1.38, p=0.82) and 24-month PR of 0.83 (95% CI: 0.53, 1.28; p=0.39) (Table 2).

The baseline mean perceived stress score was 12.6 (Table 1). Changes in mean stress scores, which slightly decreased for each cohort over time, are depicted in Table S4. There was no evidence of an intervention benefit on mean stress scores. The 12-month MD was 0.10 (95% CI: -0.38, 0.58; p=0.67); the 24-month MD was -0.28 (95% CI: -0.98, 0.42; p=0.44) (Table 2).

Table 1. Baseline characteristics of the Spirited Life study participants by randomized cohort (N=1,114)

Characteristics	Immediate Intervention Cohort (N=395)	One-year Waitlist Cohort (N=283)	Two-year Waitlist Cohort (N=436)	Total (N=1,114)	Sample Size
	No. (%)				
Male gender	271 (68.6)	199 (70.3)	302 (69.3)	772 (69.3)	1,114
Race					1,114
White	352 (89.1)	250 (88.3)	389 (89.2)	991 (89.0)	
African American	25 (6.3)	18 (6.4)	25 (5.7)	68 (6.1)	
Other	18 (4.6)	15 (5.3)	22 (5.0)	55 (4.9)	
Highest level of education achieved					1,112
College and below	61 (15.5)	47 (16.6)	85 (19.5)	193 (17.4)	
Master's	285 (72.5)	201 (71.0)	302 (69.3)	788 (70.9)	
Doctorate	47 (12.0)	35 (12.4)	49 (11.2)	131 (11.8)	
Married	353 (89.6)	252 (89.0)	386 (88.5)	991 (89.0)	1,113
Appointed to rural (vs urban) church	119 (30.3)	97 (34.3)	148 (34.3)	364 (32.8)	1,108
BMI categories					1,104
Obese	198 (50.4)	133 (47.2)	207 (48.3)	538 (48.7)	
Overweight	131 (33.3)	97 (34.4)	148 (34.5)	376 (34.1)	
Normal/underweight	64 (16.3)	52 (18.4)	74 (17.2)	190 (17.2)	
Metabolic syndrome	193 (49.2)	142 (50.9)	225 (52.3)	560 (50.9)	1,101
Central obesity	321 (82.1)	234 (83.3)	339 (79.0)	894 (81.2)	1,101
Elevated triglycerides	186 (47.4)	140 (50.4)	232 (54.3)	558 (50.9)	1,097
Low high-density lipoprotein	209 (53.5)	157 (56.5)	263 (61.7)	629 (57.4)	1,095
Hypertension	209 (53.2)	147 (52.1)	226 (52.4)	582 (52.6)	1,106
Abnormal glucose regulation	42 (10.8)	46 (16.9)	60 (14.3)	148 (13.7)	1,083
PHQ-8 depression	59 (15.1)	31 (11.0)	36 (8.4)	126 (11.4)	1,104
	Mean (SD)				
Age, yr	51.6 (10.0)	51.7 (10.1)	52.3 (9.9)	51.9 (10.0)	1,107
Weight, kg	95.3 (23.7)	93.6 (23.4)	94.6 (23.4)	94.6 (23.5)	1,103
Body mass index, kg/m^2	31.6 (7.4)	30.9 (7.2)	31.0 (7.1)	31.2 (7.3)	1,103
PHQ-8 depressive symptoms	4.6 (4.5)	4.1 (4.1)	3.9 (3.7)	4.2 (4.1)	1,104
Perceived stress	13.0 (6.3)	12.5 (6.1)	12.4 (6.1)	12.6 (6.2)	1,100

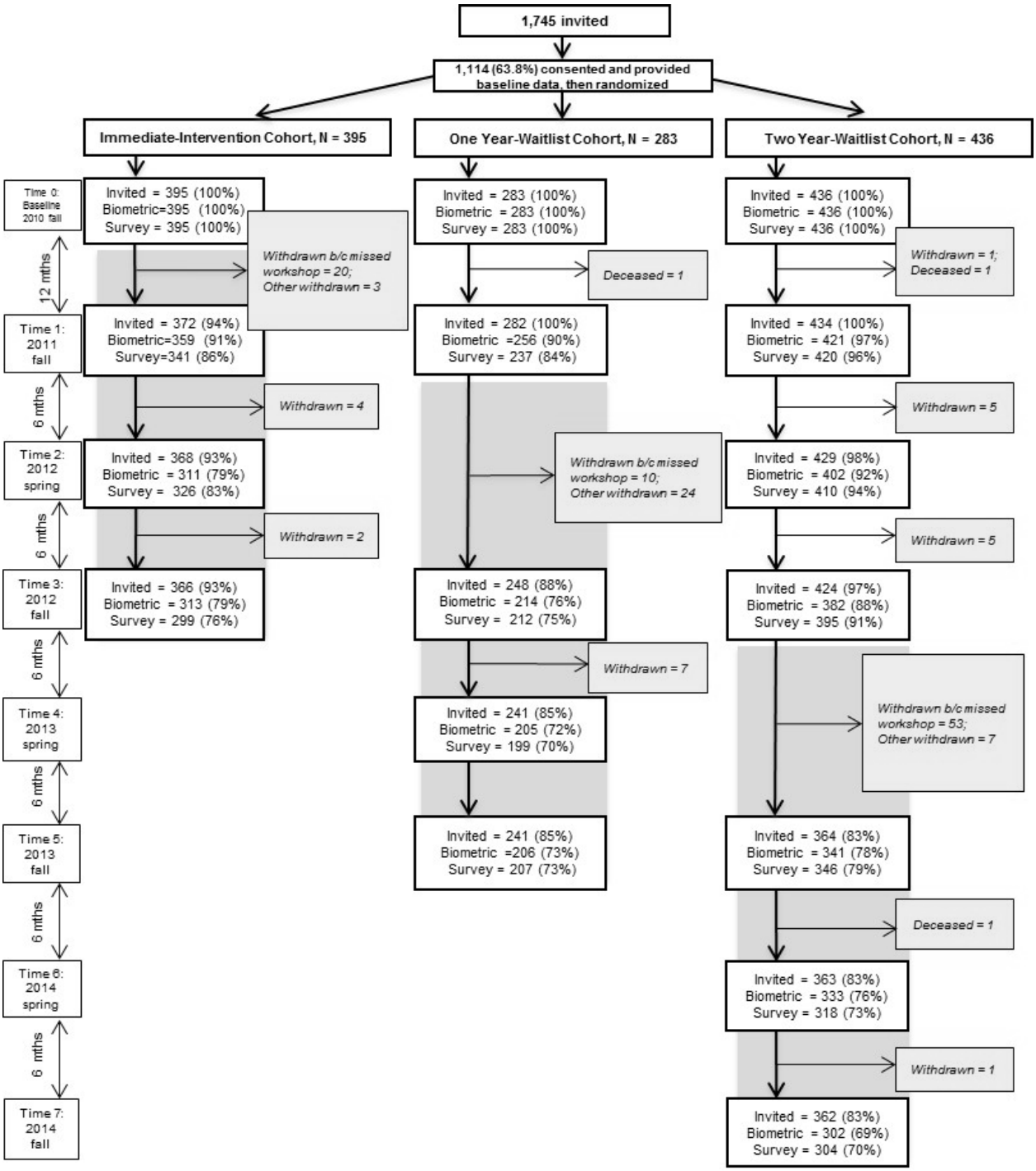
Note: Eight Spirited Life participants were pregnant or within six months postpartum at baseline. Therefore, they were excluded for the metrics of weight, BMI, metabolic syndrome, central obesity, elevated triglycerides, low high-density lipoprotein, hypertension, abnormal glucose regulation, and depression.

Table 2. Effectiveness of the Spirited Life intervention on main health outcomes by intervention duration (N=1,054)

	12-Month Intervention Effect	18-Month Intervention Effect	24-Month Intervention Effect
Prevalence Ratios (95% CI); p value			
Metabolic syndrome	**0.86 (0.79, 0.94)***; p<0.001**	**0.78 (0.69, 0.90)***; p<0.001**	**0.88 (0.78, 1.00)*; p=0.042**
Central obesity	**0.93 (0.89, 0.97)***; p<0.001**	**0.92 (0.87, 0.97)**; p=0.003**	**0.91 (0.86, 0.96)***; p<0.001**
Elevated triglycerides	**0.87 (0.79, 0.96)**; p=0.005**	**0.83 (0.71, 0.97)*; p=0.020**	0.96 (0.84, 1.09); p=0.532
Low high-density lipoprotein	**0.92 (0.86, 0.98)*; p=0.016**	**0.86 (0.78, 0.95)**; p=0.003**	**0.90 (0.81, 1.00)*; p=0.041**
Hypertension	**0.80 (0.74, 0.87)***; p<0.001**	**0.85 (0.75, 0.96)*; p=0.010**	**0.81 (0.72, 0.91)***; p<0.001**
Abnormal glucose regulation	1.00 (0.88, 1.14); p=0.961	1.10 (0.91, 1.32); p=0.347	0.98 (0.81, 1.20); p=0.876
PHQ-8 depression	1.03 (0.78, 1.38); p=0.818	0.94 (0.62, 1.44); p=0.790	0.83 (0.53, 1.28); p=0.389
Mean Differences (95% CI); p value			
Perceived stress	0.10 (-0.38, 0.58); p=0.67	0.44 (-0.22, 1.11); p=0.19	-0.28 (-0.98, 0.42); p=0.44

Notes: For each intervention level (12 months, 18 months, or 24 months in intervention vs no intervention), prevalence ratios are estimated for binary outcomes (metabolic syndrome, components of metabolic syndrome, and depression) using Poisson GEE and mean differences are estimated for the score outcome (stress), using Gaussian GEE regression modeling, All models adjust for time, district, and the baseline measure of the respective outcome and use an unstructured working correlation matrix and robust standard errors (to account for outcome misspecification). Ninety-five percent confidence intervals are reported. Boldface indicates statistical significance (*p<0.05, **p<0.01, ***p<0.001). For abnormal glucose regulation and depression, the correlation structure of the model is specified as exchangeable to avoid convergence problems. For all outcomes except perceived stress, data were collected and analyzed through 48 months from baseline. For perceived stress only, data were not collected at 42 and 48 months and therefore those time points were not included in the perceived stress analysis.

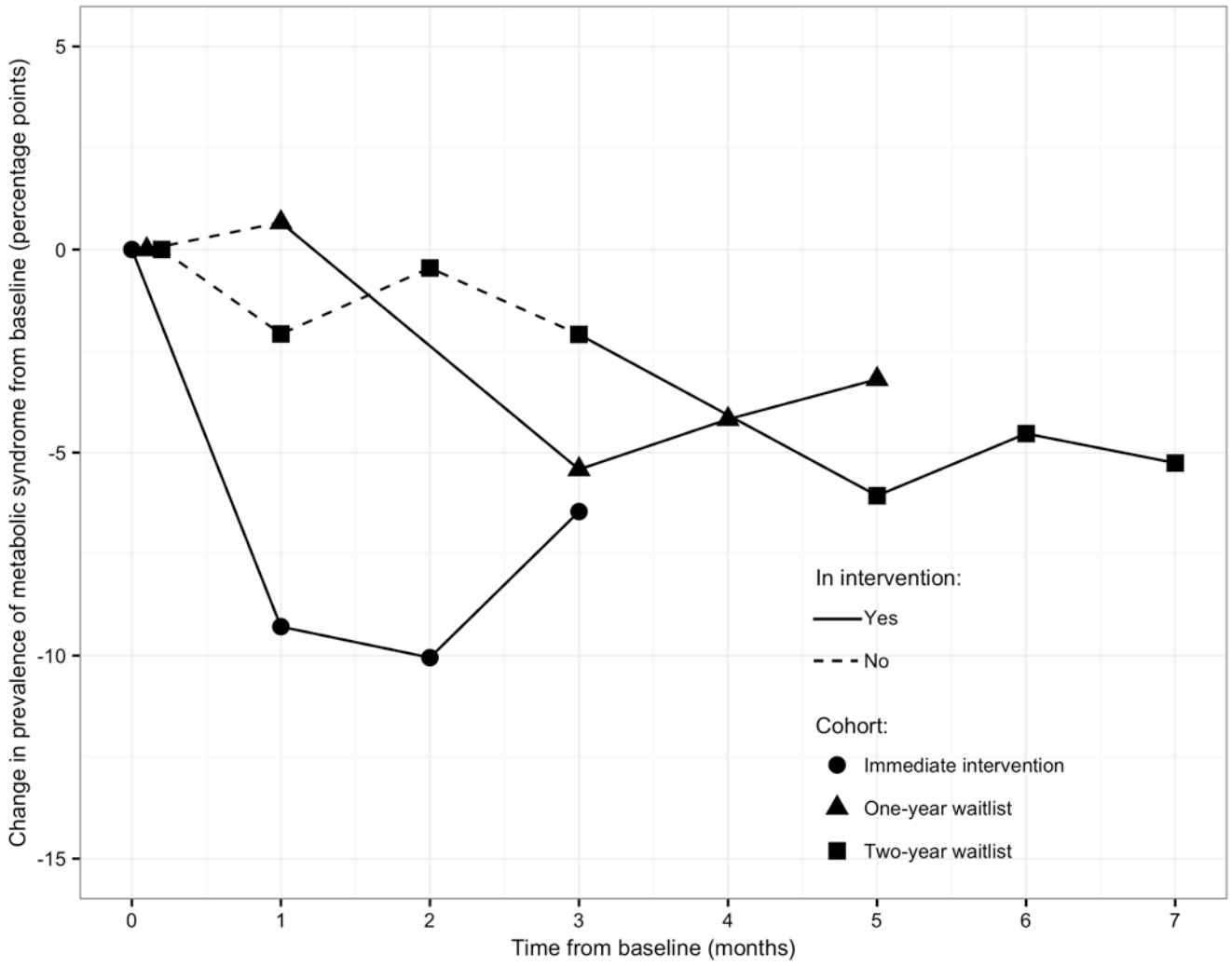
Figure 1. Trial profile: Cardiometabolic (biometric) and survey data collection at each assessment point during the 48 month study.



Legend: Shading indicates when the Spirited Life intervention was delivered to each cohort (with start dates each January, spaced 1 year apart).

Note: Percentages of cardiometabolic assessment participants and of survey participants were calculated with the number of subjects randomized as a denominator.

Figure 2. Change in the prevalence of metabolic syndrome over time by intervention cohort.



Note: The estimated prevalence changes are based on imputed data (see Supplement). Each change score is calculated as the cohort's estimated prevalence of metabolic syndrome (MetS) at a given follow-up time, minus the prevalence of MetS at time 0. For image clarity, the baseline prevalence scores are slightly shifted on the time axis, but all were measured at time 0.

Discussion

The Spirited Life trial demonstrates that a two-year intervention providing culturally-tailored content supporting healthy behaviors and training in stress management and weight loss can improve and, importantly, sustain changes during 24 months of intervention in MetS, central obesity, HDL, and hypertension at a population-level. Although few interventions have such a long duration, participants were willing to engage in a two-year intervention. The long duration may have allowed participants to practice healthy behaviors and still have support in place to minimize any lapses in those behaviors. The inclusion of a small grant at the intervention midpoint may have encouraged continued participation and assisted with maintaining newly established healthy behaviors.

The primary aim was to decrease MetS prevalence in this high-risk population, and at 12 months of intervention, there was a 14% lower prevalence of MetS. Improvements in MetS prevalence were maintained over time with a 22% lower prevalence at 18 months and 12% lower prevalence at 24 months of intervention. Of the five components of MetS, the beneficial effects of the intervention at 24 months were mainly driven by prevalence improvements in central obesity, HDL cholesterol, and blood pressure. Benefits to the prevalence of elevated triglycerides were observed at 12 and 18 months only, despite prevalence improvements in central obesity. Benefits were not observed for abnormal glucose regulation (HbA1c), which was the least prevalent MetS component in the sample and did not show a trend toward improvement. Future interventionists targeting MetS should consider including a specific diabetes program component.

Because Spirited Life took a population-level approach and invited all UMC clergy in North Carolina regardless of health status or readiness for behavioral change, enrollment was likely de-stigmatized. Sixty-four percent of invited clergy enrolled. Many participants reported they enrolled to be supportive of other clergy in their Conference, rather than wanting to change their own behavior. Participants were not required to engage in any specific intervention activity other than the initial three-day workshop, nor did they have to focus their energy on a metabolic outcome if they preferred to pursue other goals (e.g., spiritual well-being). However, after enrollment many participants engaged in multiple intervention activities (see Supplement), suggesting that interventionists should focus efforts on initial enrollment and culturally tailoring programming for maximum acceptability. Another advantage of this population-level approach was the possibility of broadly and positively influencing social norms.

The disadvantages of this population approach included spending intervention resources on participants without current health needs and likely attenuation of key outcomes. For example, not every participant was obese or motivated to lose weight. With an estimated 24-month intervention effect of -1.75 kg compared to control participants, the weight change observed in this trial was less than those that employ obesity inclusion criteria and participant interest in losing weight.[35,36] Nevertheless, at 24 months, 47.3% of the immediate-intervention cohort lost 3% or more of their baseline body weight, a percentage which obesity treatment guidelines indicate can produce clinically meaningful reductions in triglycerides and blood glucose.[29] By comparison, in a YMCA-effectiveness study of the six-month Diabetes Prevention Program (DPP) plus an eight-month maintenance intervention, participants sustained an average loss of 4.8% of their baseline weight at 28 months.[37] The DPP is a much more intensive lifestyle intervention than Spirited Life, making the current study's findings notable.[38] To scale Spirited Life in the future, it may be possible to exchange its health coaching for the recent DPP scaling work that uses online health coaching and peer groups (https://www.omadahealth.com/solution), especially if peer groups of clergy could be formed. The long

intervention periods of Spirited Life and the DPP may be key to sustaining weight loss; weight gain is common in the absence of weight maintenance programming.[40]

Random imbalance in depression was observed at baseline; the immediate-intervention cohort started with a higher prevalence than the other cohorts, which was accounted for in modelling by baseline adjustment. The intervention did not have a significant effect on depression at 12, 18, or 24 months of intervention. The lack of impact on depression prevalence could be due to the intervention's focus on stress rather than depression, although clinical trials of WLS demonstrated reductions in depression levels among patients post-coronary bypass surgery and caregivers for relatives with Alzheimer's disease.[41,42] Null findings could also be due to an ineffectual intervention for depression, the difficulty of reducing prevalence in low-prevalent disorders, and/or less severe depression in this sample.

A study hypothesis was that weight loss would be better sustained in the presence of improvements in stress symptoms. However, baseline mean stress scores were lower than the literature indicating a large number of stressors for clergy would suggest.[9,43,44] Study authors investigated this discrepancy by conducting a cognitive interviewing study with 12 clergy for each item on the PSS-10 and found that at least half had theological concerns with all but three items.[45] Clergy indicated that items such as "things are going your way" and "you could not overcome" directly conflict with seeking God's way and being faithful. Because of these concerns, this study's changes in stress scores cannot be interpreted. It is difficult to know whether: this measure is invalid for clergy; floor effects limited the possibility of finding a change in scores; or there was no true impact of this individual-level intervention on perceived stress, given the systems-level stressors experienced by UM clergy (e.g., complex church dynamics, a shrinking denomination). Researchers should explore other ways to measure perceived stress in clergy, such as the Clergy Occupational Distress Index.[46] They should continue to seek interventions that decrease stress and depressive symptoms among clergy, which may affect weight loss and are important in their own right. Given the modest (e.g., 1.7kg) weight loss findings, it does not appear that the combined stress management-weight loss intervention resulted in greater weight loss than weight-loss interventions alone. However, as noted earlier, this could be because participants were recruited based on clergy rather than obese status.

Limitations

Trial limitations include the use of self-report measures for stress and depression and power to detect only large effect sizes in those outcomes. Using waitlist control groups with clergy who regularly interact may have resulted in spillover effects; if so, outcomes may be under-estimated. One study strength was its attention to religious culture through including theological reasons to attend to health. However, this may confine the generalizability of study findings primarily to U.S. Christian clergy, although with minor adaptations, Spirited Life may be extended to the large church-affiliated population in the U.S. Other study strengths include the collection of cardiometabolic data, a large sample size, a long intervention duration, and use of a randomized multiple baseline design.

Conclusion

This trial demonstrates that the Spirited Life intervention is beneficial to U.S. Christian clergy in improving MetS, central obesity, HDL, and hypertension, as well as sustaining these improvements during 24 months of intervention. These findings offer support for long-duration behavior change interventions and

population-level interventions that allow participants to set their own health goals. Future studies should continue to test interventions aimed at the dual goals of MetS and stress symptom reduction powered to detect meaningful but smaller stress and depression reductions than targeted here, and should consider testing multi-year weight loss programs with an eye toward enhanced scalability.

Acknowledgements

The United Methodist clergy of North Carolina, along with support from their bishops and leaders, made this trial possible. Data were collected and organized by Crystal MacAllum, Gail Thomas, Ed Mann, and their team at Westat, as well as Amanda Wallace and Alana Bennett. Thirteen Wellness Advocates implemented the intervention with great care. Amanda Wallace, and Adeola Awodele helped prepare the manuscript. This trial was funded by the Rural Church Area of The Duke Endowment, http://dukeendowment.org/. The Duke University Arts & Sciences IRB protocol number is A0208. The tasks completed by each author are as follows: Conceived and designed the trial: RJPB GGB RW DT HEM RS; Performed the trial (acquired the data): RJPB RS HEM MK RM CW KR HH DT. Analyzed the data: ELT XFL JY DE. Interpreted the data: RJPB ELT GGB XFL JY RW VW DT KR RM CW HEM RS DE. Wrote the first draft of the manuscript: RJPB ELT GGB RW. Provided critical revision of the manuscript for important intellectual content: RJPB ELT GGB XFL JY RW RS HEM MK RM CW KR VW DT DE. Obtained funding: DT. Administrative, technical or material support: GGB XFL JY RW RS HEM MK RM CW KR VW. Supervision: RJPB ELT CW DT RS. The trial registration number in www.ClinicalTrials.gov is NCT01564719. Trial rationale but not outcomes were presented at the 34th Annual Meeting & Scientific Sessions of the Society of Behavioral Medicine, the APA Division 27 2013 Biennial Conference, and the Society for the Scientific Study of Religion 2011 Annual Meeting. Financial disclosures: Gary G.Bennett holds equity in Scale Down and Coeus Health, which produce digital health interventions for obesity. He also serves on the science advisory board at Nutrisystem and the board of directors at Girl Trek. Redford B. Williams is a founder of and major stockholder in Williams LifeSkills, Inc. Virginia P. Williams is a founder of and major stockholder in Williams LifeSkills, Inc.

Statement for Replication and Review

The data used for this study may be obtained from Rae Jean Proeschold-Bell, first author, for purposes of replication. She is best contacted over email at rae.jean@duke.edu. This research was funded by a grant from The Duke Endowment.

Conflict of Interest Statement

Rae Jean Proeschold-Bell has no conflicts. Her work on this paper was funded by The Duke Endowment.
Elizabeth L. Turner has no conflicts. Her work on this paper was funded by The Duke Endowment.
Xiang-Fang Li has no conflicts. Her work on this paper was funded by The Duke Endowment.
Jia Yao has no conflicts. Her work on this paper was funded by The Duke Endowment.
Gary G. Bennett has no conflicts. He was not funded to work on this paper.
David E. Eagle has no conflicts. His work on this paper was funded by The Duke Endowment.
Rachel A. Meyer has no conflicts. Her work on this paper was funded by The Duke Endowment.
Redford B. Williams has no conflicts. He was not funded to work on this paper.
Robin Y. Swift has no conflicts. Her work on this paper was funded by The Duke Endowment.
H. Edgar Moore has no conflicts. His work on this paper was funded by The Duke Endowment.

Melanie A. Kolkin has no conflicts. Her work on this paper was funded by The Duke Endowment.
Carl C. Weisner has no conflicts. His work on this paper was funded by The Duke Endowment.
Katherine M. Rugani has no conflicts. Her work on this paper was funded by The Duke Endowment.
Holly J. Hough has no conflicts. Her work on this paper was funded by The Duke Endowment.
Virginia P. Williams has no conflicts. Her work on this paper was funded by The Duke Endowment.
David C. Toole has no conflicts. His work on this paper was funded by The Duke Endowment.

The study sponsor, The Duke Endowment, had no role in study design; collection, analysis, or interpretation of data; writing the report; or final decision to submit the report for publication.

Financial disclosure statement

Rae Jean Proeschold-Bell has no financial disclosures.
Elizabeth L. Turner has no financial disclosures.
Xiang-Fang Li has no financial disclosures.
Jia Yao has no financial disclosures.
Gary G. Bennett holds equity in Scale Down and Coeus Health, which produce digital health interventions for obesity. He also serves on the science advisory board at Nutrisystem and the board of directors at Girl Trek.
David E. Eagle has no financial disclosures.
Rachel A. Meyer has no financial disclosures.
Redford B. Williams is a founder of and major stockholder in Williams LifeSkills, Inc.
Robin Y Swift has no financial disclosures.
H. Edgar Moore has no financial disclosures.
Melanie A. Kolkin has no financial disclosures.
Carl C. Weisner has no financial disclosures.
Katherine M. Rugani has no financial disclosures.
Holly J. Hough has no financial disclosures.
Virginia P. Williams is a founder of and major stockholder in Williams LifeSkills, Inc.
David C. Toole has no financial disclosures.

Trial registration www.ClinicalTrials.gov: NCT01564719; date of registration: 3/14/2012

Trial funder: The Duke Endowment

Institutional Review Board: Duke University Arts & Sciences IRB, #A0208; Date of approval: September 24, 2010

References

1. Alberti KGMM, Zimmet P, Shaw J. Metabolic syndrome—a new world-wide definition. A Consensus Statement from the International Diabetes Federation. Diabet Med. 2006;23(5):469-480.
2. Lorenzo C, Okoloise M, Williams K, Stern MP, Haffner SM. The metabolic syndrome as predictor of type 2 diabetes the San Antonio heart study. Diabetes Care. 2003;26(11):3153-3159.
3. Isomaa B, Almgren P, Tuomi T, et al. Cardiovascular morbidity and mortality associated with the metabolic syndrome. Diabetes Care. 2001;24(4):683-689.
4. Jeffery RW, Epstein LH, Wilson GT, Drewnowski A, Stunkard AJ, Wing RR. Long-term maintenance of weight loss: current status. Health Psychol. 2000;19(1S):5.
5. Dansinger ML, Tatsioni A, Wong JB, Chung M, Balk EM. Meta-analysis: the effect of dietary counseling for weight loss. Ann Intern Med. 2007;147(1):41-50.
6. Dallman MF, Pecoraro NC, la Fleur SE. Chronic stress and comfort foods: self-medication and abdominal obesity. Brain Behav Immun. 2005;19(4):275-280.
7. Cox TL, Krukowski R, Love SJ, et al. Stress Management–Augmented Behavioral Weight Loss Intervention for African American Women A Pilot, Randomized Controlled Trial. Health Educ Behav. 2013;40(1):78-87.
8. Daubenmier J, Kristeller J, Hecht FM, et al. Mindfulness intervention for stress eating to reduce cortisol and abdominal fat among overweight and obese women: an exploratory randomized controlled study. J Obes. 2011;2011 (Article ID 651936):35.
9. Lee C, Iverson-Gilbert J. Demand, support, and perception in family-related stress among Protestant clergy. Fam Relat. 2003;52(3):249-257.
10. Knox S, Virginia SG, Thull J, Lombardo JP. Depression and contributors to vocational satisfaction in Roman Catholic secular clergy. Pastoral Psychol. 2005;54(2):139-155.
11. Proeschold-Bell RJ, Miles A, Toth M, Adams C, Smith BW, Toole D. Using effort-reward imbalance theory to understand high rates of depression and anxiety among clergy. J Prim Prev. 2013;34(6):439-453.
12. General Board of Pensions and Health Benefits of the United Methodist Church. Annual clergy health survey: Report. 2013; http://www.gbophb.org/assets/1/7/2013_Clergy_Health_Survey_Report.pdf. Accessed August 15, 2016.
13. Proeschold-Bell RJ, LeGrand SH. High rates of obesity and chronic disease among United Methodist clergy. Obesity. 2010;18(9):1867-1870.
14. The General Council on Finance and Administration of the United Methodist Church. The state of female and racial/ethnic United Methodist clergy in the US. 2012; http://www.umc.org/gcfa/the-state-of-female-and-racial-ethnic-united-methodist-clergy-in-the-u.s. Accessed July 27, 2016.
15. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity in the United States, 2009–2010. NCHS Data Brief. 2012;82:1-8.
16. Befort C. Bridging the gender gap in behavioral obesity intervention: A comment on Morgan et al. Ann Behav Med. 2013;45(2):133-134.
17. U.S. Department of Labor. Occupational Outlook Handbook, 2014 Edition. 2015; http://www.bls.gov/ooh/about/data-for-occupations-not-covered-in-detail.htm. Accessed August 24, 2016.
18. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol. 2009;62:464-475.

19. Koppenaal T, Linmans J, Knottnerus JA, Spigt M. Pragmatic vs. explanatory: an adaptation of the PRECIS tool helps to judge the applicability of systematic reviews for daily practice. J Clin Epidemiol. 2011;64(10):1095-1101.
20. Proeschold-Bell RJ, Swift R, Moore HE, et al. Use of a randomized multiple baseline design: Rationale and design of the Spirited Life holistic health intervention study. Contemp Clin Trials. 2013;35(2):138-152.
21. Williams RB, Williams VP. In Control: No More Snapping at Your Family, Sulking at Work, Steaming in the Grocery Line, Seething in Meetings, Stuffing your Frustration. New York: Rodale Books; 2006.
22. NHLBI Obesity Education Initiative Expert Panel on the Identification, Evaluation, and Treatment of Obesity in Adults (US). Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report. Bethesda, MD: National Heart, Lung, and Blood Institutes; 1998.
23. Spitzer RL, Kroenke K, Williams JB, The Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. JAMA. 1999;282(18):1737-1744.
24. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord. 2009;114(1):163-173.
25. Zeger SL, Liang K-Y, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics. 1988:1049-1060.
26. Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks from clustered prospective data. Am J Epidemiol. 2011:kwr183.
27. Handley MA, Schillinger D, Shiboski S. Quasi-experimental designs in practice-based research settings: design and implementation considerations. J Am Board Fam Med. 2011;24(5):589-596.
28. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. Stat Med. 1992;11(13):1685-1704.
29. Jensen MD, Ryan DH, Apovian CM, et al. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. J Am Coll Cardiol. 2014;63(25_PA).
30. Blackburn G. Effect of degree of weight loss on health benefits. Obes Res. 1995;3(S2):211s-216s.
31. Van Gaal LF, Mertens IL, Ballaux D. What is the relationship between risk factor reduction and degree of weight loss? Eur Heart J Suppl. 2005;7(suppl L):L21-L26.
32. Wing RR, Lang W, Wadden TA, et al. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. Diabetes Care. 2011;34(7):1481-1486.
33. Little RJA, Rubin DB. Statistical Analysis with Missing Data, 2nd Edition. Wiley-Interscience; 2002.
34. Jolani S, van Buuren S. Doubly Robust Imputation of Incomplete Binary Longitudinal Data. J Biom Biostat. 2014;2014.
35. Franz MJ, VanWormer JJ, Crain AL, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J Am Diet Assoc. 2007;107(10):1755-1767.
36. Curioni C, Lourenco P. Long-term weight loss after diet and exercise: a systematic review. Int J Obes. 2005;29(10):1168-1174.
37. Ackermann RT, Finch EA, Caffrey HM, Lipscomb ER, Hays LM, Saha C. Long-term effects of a community-based lifestyle intervention to prevent type 2 diabetes: the DEPLOY extension pilot study. Chronic Illn. 2011:1742395311407532.
38. Group DPPR. The Diabetes Prevention Program (DPP) description of lifestyle intervention. Diabetes Care. 2002;25(12):2165-2171.

39. Wing RR, Koeske R, Epstein LH, Nowalk MP, Gooding W, Becker D. Long-term effects of modest weight loss in type II diabetic patients. Arch Intern Med. 1987;147(10):1749-1753.
40. Butryn ML, Webb V, Wadden TA. Behavioral treatment of obesity. Psychiatr Clin North Am. 2011;34(4):841.
41. Bishop GD, Kaur D, Tan VL, Chua Y-L, Liew S-M, Mak K-H. Effects of a psychosocial skills training workshop on psychophysiological and psychosocial risk in patients undergoing coronary artery bypass grafting. Am Heart J. 2005;150(3):602-609.
42. Williams VP, Bishop-Fitzpatrick L, Lane JD, et al. Video-Based Coping Skills (VCS) to Reduce Health Risk and Improve Psychological and Physical Well-being in Alzheimer's Disease Family Caregivers. Psychosom Med. 2010;72(9):897.
43. Proeschold-Bell RJ, LeGrand S, James J, Wallace A, Adams C, Toole D. A theoretical model of the holistic health of United Methodist clergy. J Relig Health. 2011;50(3):700-720.
44. Morris ML, Blanton P. Predictors of family functioning among clergy and spouses: Influences of social context and perceptions of work-related stressors.
J Child Fam Stud. 1998;7(1):27-41.
45. Blouin R, Proeschold-Bell RJ. Measuring Stress in a Clergy Population: Lessons Learned from Cognitive Interview Testing of the Perceived Stress Scale with Clergy. Res Soc Sci Study Relig: Brill; 2015:141-154.
46. Frenk SM, Mustillo SA, Hooten EG, Meador KG. The Clergy Occupational Distress Index (CODI): Background and findings from two samples of clergy. J Relig Health. 2013;52(2):397-407.

Online-Only Supplemental Material

for

Impact of a Two-Year Multi-Component Health and Stress Management Intervention for Clergy: A Randomized Controlled Trial

by

Rae Jean Proeschold-Bell, Elizabeth L. Turner, Gary G. Bennett, Jia Yao, Xiang-Fang Li, David E. Eagle, Rachel A. Meyer, Redford B. Williams, Robin Y. Swift, H. Edgar Moore, Melanie A. Kolkin, Carl C. Weisner, Kathrine M. Rugani, Holly J. Hough, Virginia P. Williams, and David C. Toole

This document contains the following sections:

Supplemental Information Section 1: Supplemental Information on Methods

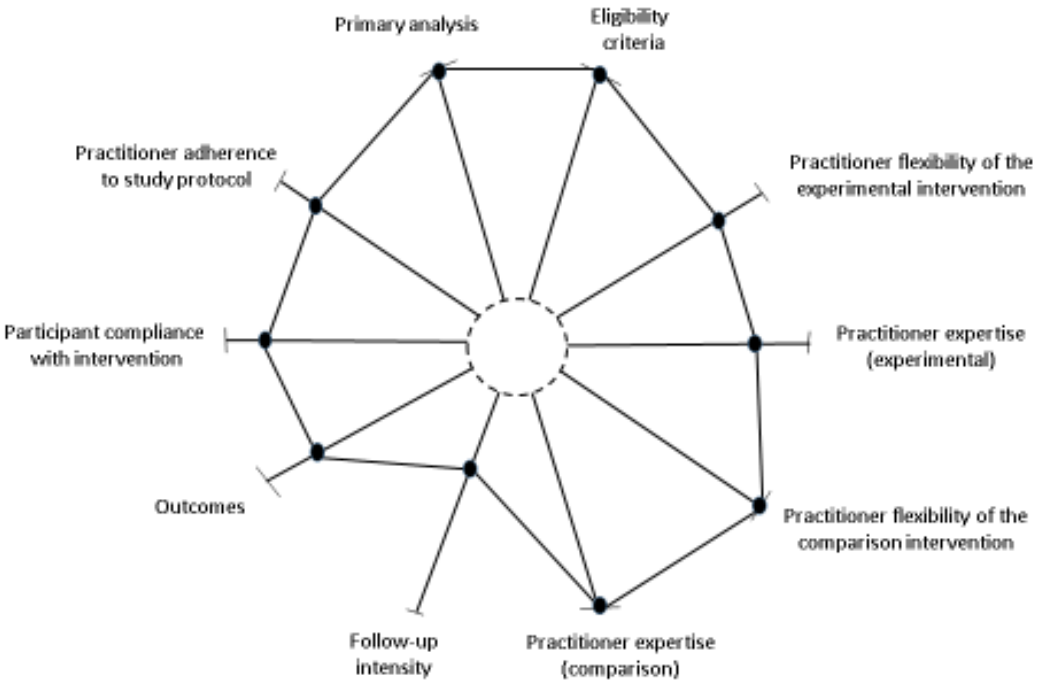
Intervention details

Spirited Life was conducted as a pragmatic trial. Using pragmatic-explanatory continuum indictor summary (PRECIS) criteria,[1] we consider Spirited Life to be highly pragmatic, with an estimated score of 85% out of 100%, with 100% being extremely pragmatic.[2] As can be seen in the PRECIS Supplement Figure 1, we rate four of the ten criteria as extremely pragmatic. All clergy were eligible, with the only exclusion criteria being specific clergy statuses and not based on health. For the comparison intervention, we used a waitlist control condition in which we did not limit any participant health-seeking activities. Relatedly, the practitioners for the comparison intervention, therefore, had typical levels of expertise. Our analysis of the primary outcome was an intent-to-treat analysis which is the most pragmatic kind of analysis.

We rated four criteria as "very" rather than "extremely" pragmatic. Our outcomes have real-world significance and do not require central adjudication, although the measurement of cholesterol and hemoglobin A_{1c} require special equipment. The practitioners who carried out the intervention included Wellness Advocates, whom we certified in health coaching and Williams LifeSkills. In terms of auditing practitioner adherence to the study protocol, we did not audiotape or listen into health coaching sessions. However, we did ask Wellness Advocates to conduct role plays with the health coach expert and the Program Director. These same practitioners were given wide berth in implementing health coaching, but were instructed to reach out to non-responsive participants at least three times. Participants were allowed to participate in just the intervention components they desired, but we required that all participants attend the initial three-day workshop.

We were least pragmatic with the intensity of our follow-up data collection. We collected data twice during the intervention period, as well as just prior and just after. Participants would not have had these data collected (e.g., in a routine physical) if they had not been in the trial.

Supplement Figure 1. PRECIS criteria and scores for the Spirited Life trial, with each criteria having five points with most pragmatic being furthest from the center.



Intervention health coaches contacted participants after their initial workshop to set up monthly health coaching calls. If participants did not schedule a call, their health coach made monthly attempts to contact participants for three months, and then sent occasional emails with general health information and intervention updates to remind participants of their availability. The health coaches came from a variety of educational backgrounds and were trained prior to study launch by experts in Motivational Interviewing and Williams LifeSkills (WLS). Health coaches' skills on Motivational Interviewing were assessed monthly, and all health coaches met weekly to discuss intervention implementation.

Cardiometabolic data collection

Cardiometabolic data collection assessed the five components of metabolic syndrome and was performed by staff trained on detailed protocols (available upon request). In brief, participants were asked to fast for at least 9 hours prior to data collection. Finger sticks were used to collect blood samples for lipid tests and glycated hemoglobin (HbA_{1c}). Lipid values were measured by the Cholestech LDX system. The lipids of interest for metabolic syndrome are high-density lipoprotein (HDL) and triglycerides. HbA1c was analyzed using the Afinion HbA1c test. After an initial five-minute resting period, blood pressure was assessed three times on the right arm using an Omron HEM-907XL machine, with a thirty second rest between measurements. During the rest and measurement periods, participants remained seated with feet flat on the floor and arm resting at heart level. The mean of the

three values was used in data analysis. Waist circumference was defined as the abdominal girth halfway between the iliac crest and the lower costal margin and measured to the nearest 0.25 inch. In addition to these metabolic syndrome indicators, we measured height to the nearest 0.25 inch using a Seca 213 stadiometer, and weight to the nearest 0.2 lb. on a high-quality, calibrated digital scale (Seca 876), with the participant removing heavy clothing and shoes prior to being weighed.

Cardiometabolic data were recorded by assessor staff on paper and double data-entered. Consistency checks were performed on all data sets in order to create reliable final databases for analysis.

Metabolic syndrome definition

Prior to the trial, we pre-specified that we would use the definition of the International Diabetes Federation.[3] In this definition, central obesity is defined as a waist circumference greater than or equal to 94 cm for men and 80 cm for women. Raised triglycerides are defined as greater than or equal to 150 mg/dl. Reduced HDL-cholesterol is defined as less than 40 mg/dl in men and less than 50 mg/dl in women. Raised blood pressure is defined as systolic blood pressure greater than or equal to 130 mmHg or diastolic blood pressure greater than or equal to 85 mmHg. The International Diabetes Federation defines diabetic risk in terms of fasting plasma glucose of greater than or equal to 5.6 mmol/l. However, we measured diabetic risk using HbA_{1c} rather than fasting plasma glucose levels. We did this because HbA_{1c} is not sensitive to fasting, provides a longer-term measure of degree of glucose exposure measure, and is more closely related to the risk of health complications than single measures of glucose levels.[4] Following the International Expert Committee with members appointed by the American Diabetes Association, the European Association for the Study of Diabetes, and the International Diabetes Federation 2, we defined raised HbA_{1c} as greater than or equal to 6.5%. Using the metabolic syndrome criteria harmonized across organizations, we counted anyone taking medication (e.g., hypertension medication) for one of the metabolic syndrome risk factors, as qualifying for that risk factor (e.g., hypertension).[5]

Pre-specified outcomes and comparisons of interest

The pre-specified primary outcome measure was the prevalence of metabolic syndrome. The pre-specified primary comparison of interest was the two-year intervention effect estimated by comparing the immediate-intervention and the two-year waitlist cohort at 24-month follow-up (T3), after the immediate-intervention cohort had received the full two-year Spirited Life intervention and the two-year waitlist control cohort had not yet started the intervention. The pre-specified secondary outcomes were the prevalence of depression (PHQ-8 ≥10) and mean perceived stress scores, for the same comparison between the immediate-intervention and the control cohort at 24-month follow-up (T3). Weight, while not pre-specified, was also a secondary outcome. Secondary comparisons were the short- and medium-term intervention effects, which correspond to comparisons at 12-month (T1) and 18-month (T2) follow-up after the intervention was introduced in the immediate-intervention cohort. Such comparisons correspond to the effect of 12 and 18 months of intervention, with the caveat that participants began treatment across a three-month window, such that the 12-month follow-up time point actually reflects 9 to 12 months of intervention received.

Statistical power calculations

During planning, we had no prevalence data for the primary outcome but had access to chronic disease data in the NC UMC clergy population that showed that 41.2% and 36.2% were estimated to be obese or

hypertensive, respectively.[6] We estimated MetS prevalence to be 40% in the absence of intervention. With 400 clergy in each of the immediate-intervention and two-year waitlist cohorts, we would have 83% power to detect a 10 percentage point difference (40% vs. 30%). Power using the same sample size to assess secondary outcomes at 24-months was good, and exceeded 78% for all outcomes. For the prevalence of depressive symptoms, power was 78% to detect a meaningful difference of 5.6 percentage points (11.1% control vs. 5.5% immediate-intervention, based on data for clergy versus the general NC population.[7] For stress scores, power was 99% to detect a 2-point difference in mean score (13 in control vs. 11 in immediate-intervention, for a SD of 6.5 points in each group). For weight, power was 87% to detect a difference between groups of 3 kg mean weight loss over 24 months (3 kg weight loss in immediate-intervention vs. 0 kg weight loss in the control cohort, for an assumed SD of weight loss of 13.7 kg).

Supplemental Information Section 2: Supplemental Information on Results

Missing data analysis

In this section of the supplementary material, we describe the comprehensive approach we used to perform a missing outcome data sensitivity analysis for our primary outcome, Metabolic Syndrome (MetS). The following analysis describes which baseline covariates are predictive of missing outcome variables; patterns of missing MetS by follow-up time point; under what circumstances Generalized Estimating Equations (GEE) are valid; and the multiple imputation method we used in our sensitivity analysis. Overall, the sensitivity analyses suggest that the main results are robust to missing data.

Predictors of missing metabolic syndrome measurements. In order to test if participants with particular characteristics were more likely to drop out, the baseline characteristics of participants who remained in the study versus those who dropped out were compared. The focus was on eight different baseline characteristics that could potentially predict missingness: female gender, having a college degree or below, BMI status, baseline MetS status, PHQ-8 score of 10 or more, age, and cohort membership. The main predictors of missingness to emerge were: college or lower educational attainment, treatment cohort, female gender, rural location, and baseline MetS status.

There were very few missing values for any of the baseline predictors of missingness. Out of the 1,114 study participants, 13 people had missing data that prevented calculation of baseline MetS status, 6 were missing information on rural location, 11 were missing data to calculate BMI, 10 were missing PHQ-8 scores, and 7 were missing age.

Patterns of missing by follow-up time point. Of the 1,114 study participants, there were 13 (1.3%) missing MetS values at time 1, 52 (6.8%) missing values at time 2, 74 (7.6%) missing values at time 3, 8 (3.8%) missing values at time 4, 32 (5.6%) missing values at time 5, 13 (3.8%) missing values at time 6, and 35 (10.9%) missing values at time 7. Participants with missing data at one wave did not necessarily have missing data at subsequent waves.

Methods. The Generalized Estimating Equations (GEE) approach was used to estimate the intervention effect (see Statistical Analysis section of the main manuscript).[8] The GEE approach is valid only under the missing completely at random assumption (MCAR) or under the covariate dependent missing (CDM) assumption. MCAR holds if missing values are independent of both observed and unobserved data. CDM holds if missing outcome values depend on only baseline covariates and not on any of the follow-up outcome variables.[9] In addition, in these data, there were intermittent missing values, where participants provided data at later time points after skipping one or more measurement points. Therefore, the missing data analysis approach must also deal with this feature of the data. Jolani and van Buuren (2014) propose a "doubly robust imputation method" (DR-MI) in order to impute incomplete longitudinal data.[10] This method allows the MCAR and CDM assumption to be relaxed to a missing at random (MAR) assumption. Moreover, it uses a method that is robust to misspecification in the imputation model, and that accommodates intermittent missing data. Using DR-MI, 20 datasets with imputed missing values were created. The imputation model included all of the covariates previously identified that predicted missing follow-up MetS values, as well as all previous values of MetS to predict subsequent missing MetS at later follow-up time points. GEE models with robust standard errors were estimated with each of the 20 imputed datasets and the coefficients were combined using Rubin's rules. All analyses were performed in R version 3.2.4.

Results. In Supplement Table 1, the effect of the intervention at 12 months, 18 months, and 24 months is reported for the original analysis and the DR-MI analysis. As this analysis shows, imputing these

missing values does not change the point estimates of the effect of the intervention, and only impacts the standard errors in the second decimal place for two of the effects. The fact that the coefficients and standard errors do not change (i.e. our results are robust to missing data), suggests that the results reported in the manuscript text and in Table 2 and Figure 1 are valid and appropriate to be reported as our primary results.

Supplement Table 1. Estimates of the 12, 18 and 24-month intervention effect, non-imputed and doubly robust multiple imputation models

	12-Month Intervention Effect	18-Month Intervention Effect	24-Month Intervention Effect
Prevalence Ratios (95% CI)			
Original Unimputed Model	0.86 (0.79, 0.94)	0.78 (0.69, 0.90)	0.88 (0.78, 1.00)
DR-MI Model	0.85 (0.78 0.93)	0.79 (0.69 0.90)	0.88 (0.79 0.99)

Follow-up participation

We randomized 1,114 participants. By the end of the 48-month follow-up, three participants had died and 142 (12.7%) had withdrawn (Figure 1). The majority of withdrawals occurred when participants did not attend the required first workshop, which involved travel and a three-day commitment. Specifically, 5.1% of the immediate-intervention cohort, 3.5% of the one-year waitlist cohort, and 12.2% of the two-year waitlist cohort withdrew at the time of the workshops, accounting for 58.5% of the attrition.

As expected, cohort was predictive of missing outcomes partly due to the increased attrition over time since the two-year waitlist cohort was observed over a longer period of time than the other two cohorts. Baseline outcomes indicated that, compared to the 1,054 who provided at least one follow-up measurement, the 60 (5.4%) participants with no follow-up data had poorer baseline outcomes for depression (15.0% vs. 11.2%), MetS (60.0% vs. 50.3%), and hypertension (66.7% vs. 51.8%), but not for elevated HbA1c, central obesity, elevated triglycerides, or reduced HDL. There is no evidence to suggest that the 1,054 are very different from the 1,114 enrolled participants.

There was evidence that participants with less education and those with MetS at baseline were more likely to be missing the MetS outcome. When we included education in our sensitivity models, we saw no substantive changes to our conclusions. We found no evidence to suggest that other covariates are predictive of missing outcomes at multiple time points nor that missing outcomes depend on the level of observed or unobserved outcomes once we accounted for the baseline level of each specific outcome (since outcomes measured on the same individual are correlated over time). Thus, we found no evidence to suggest that the results we report are biased as a consequence of the missing data.

Intervention participation

We report here data on the study participation of the immediate-intervention cohort. Whereas all other data reported in the manuscript use an intent-to-treat approach, the data on intervention participation reported here reflect the actual cohort assignment. Specifically, forty-two participants

originally assigned to the immediate-intervention cohort were later assigned to one of the other two cohorts, and seven participants not initially assigned to the immediate-intervention cohort, later transferred. For the immediate-intervention cohort, we offered three overnight workshops during the two-year intervention. Participation in the first workshop, which included the WLS stress management content, was required. 70.5% of participants attended all three workshops, 20.5% attended two workshops, and 9.1% attended the first workshop only. The number of health coaching sessions ranged from 0 to 20, with a mean of 6.3 (SD=4.6; median=6.0; 25^{th}, 75^{th} percentiles=3.0, 9.0). We offered ten online Naturally Slim® weight loss videos. 83.2% of the immediate-intervention cohort participants signed up for the Naturally Slim® program. Participants watched a mean number of 6.6 sessions (SD=3.8; median=8.0; 25^{th}, 75^{th} percentiles=3.0, 10.0); 11.8% did not watch any sessions, 28.2% watched between seven and nine sessions, and 34.1% watched all ten sessions. We offered $500 small grants to apply toward health goal expenditures, and 95.5% of the immediate-intervention cohort participants chose to receive these grants.

Weight outcomes (Tables S5, S6, and S7)

Supplement Table 5 shows mean weight changes by cohort for all participants (including those of normal weight at baseline) and separately for participants overweight and obese at baseline. For all participant groups, benefits were observed at all follow-up time points. After 24 months of intervention, the mean weight in the immediate-intervention cohort was 3.4 kg lower, the one-year waitlist cohort was 4.4 kg lower, and the two-year waitlist cohort was 1.7 kg lower compared to their immediately pre-intervention mean weight. In stratified analysis, we observed benefits after 24 months of intervention for participants who were obese at baseline, with a mean weight decrease of 5.1 kg in the immediate-intervention cohort, 4.9 kg in the one-year waitlist cohort, and 2.6 kg in the two-year waitlist cohort.

Supplement Table 6 displays the proportion of participants who had lost 3% and 5% of their baseline weight by cohort and time point. At 24 months, 34.1% of immediate-intervention participants had lost 5% of their baseline body weight; 19.8% of two-year waitlist participants also lost 5% of their baseline body weight prior to any intervention other that the cardiometabolic data collection. During the two years of intervention, an additional 7.6% of the two-year waitlist participants lost 5% or more of their baseline weight.

Using all intervention and control period data from all cohorts, and adjusting for follow-up time points, the 12-month intervention effect on 5% loss of baseline weight (Table S7) was estimated to be 2.3 times prevalence (PR: 2.33; 95% confidence interval [CI]: 1.95, 2.80, $p<0.001$). This effect was sustained over two years with a 24-month intervention effect estimated to be 2.0 times prevalence (PR: 2.01; 95% CI: 1.60, 2.54, $p<0.001$). The 24-month intervention benefits on the proportion of all participants achieving 3% weight loss from baseline were observed with PR of 1.70 (95% CI: 1.43, 2.02; $p<0.001$).

Supplemental Information Section 3: Supplemental Tables

Supplement Table 2. CONSORT 2010 checklist of information to include when reporting a pragmatic trial[11,12]

Section/Topic	Item No	Standard Checklist item	Extension for pragmatic trials	Section and paragraph
Title and abstract				
	1a	Identification as a randomized trial in the title		Title
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)[4, 5]		Abstract
Introduction				
Background and objectives	2a	Scientific background and explanation of rationale	Describe the health or health service problem that the intervention is intended to address; state that the trial is pragmatic; explain the purpose of the trial in relationship to the decisions that it is intended to inform and in which settings	Introduction: Paragraphs 1-5
	2b	Specific objectives or hypotheses		Introduction: Paragraphs 2, 5
Methods				
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio		Introduction: Paragraph 5, Methods: Procedure - Paragraphs 1-2, Figure 1
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons		N/A
Participants	4a	Eligibility criteria for participants	Eligibility criteria should be explicitly framed to show the degree to which they include typical participants and, where	Methods: Participants – Paragraph 1

			applicable, typical providers (e.g., nurses), institutions (e.g., hospitals), communities (or localities e.g., towns) and settings of care (e.g., different healthcare financing systems).	
	4b	Settings and locations where the data were collected		Methods: Study Population – Paragraph 1
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	Describe extra resources added to (or resources removed from) usual settings in order to implement the intervention. Indicate if efforts were made to standardize the intervention or if the intervention and its delivery were allowed to vary between participants, practitioners or study sites. Describe the comparator in similar detail to the intervention.	Methods: Intervention – Paragraph 1; Supplemental Information: Intervention Details
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	Explain why the chosen outcomes and, when relevant, the length of follow-up are considered important to those who will use the results of the trial.	Introduction: Paragraphs 1-4; Methods: Procedure – Paragraph 2; Measures – Paragraphs 1-2; Outcomes – Paragraph 1; Supplemental Information: Pre-specified Outcomes and Comparisons of Interest
	6b	Any changes to trial outcomes after the trial commenced, with reasons		N/A
Sample size	7a	How sample size was determined	If calculated using the smallest difference considered important by the target decision maker audience (the minimally important difference) then report where this difference was obtained.	Methods: Statistical Power - Paragraph 1; Supplemental Information: Statistical Power Calculations
	7b	When applicable, explanation of any interim analyses and stopping guidelines		N/A
Randomization:				

Sequence generation	8a	Method used to generate the random allocation sequence		Methods: Procedure - Paragraph 1
	8b	Type of randomization; details of any restriction (such as blocking and block size)		N/A
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned		Methods: Procedure - Paragraph 1
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions		Methods: Procedure - Paragraph 1
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	If blinding was not done, or was not possible, explain why.	Non-blinded trial Methods: Procedure - Paragraph 1
	11b	If relevant, description of the similarity of interventions		N/A
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes		Methods: Statistical Analysis – Paragraphs 1-3
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses		Methods: Statistical Analysis – Paragraph 3
Results				
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome	The number of participants or units approached to take part in the trial, the number which were eligible and reasons for non-participation should be reported.	Figure 1; Supplemental Information: Follow-up Participation
	13b	For each group, losses and exclusions after randomization, together with reasons		Figure 1; Supplemental Information: Follow-up Participation
Recruitment	14a	Dates defining the periods of recruitment and follow-up		Methods: Procedure – Paragraphs 1-2
	14b	Why the trial ended or was stopped		

				N/A
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group		Results: Sample and Follow-up Characteristics – Paragraph 1; Table 1
Numbers analyzed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups		Results: Sample and Follow-up Characteristics – Paragraphs 1-2; Figure 1
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)		Table 2: Figure 2; Supplemental Tables S4-S7 Results: MetS Outcomes Results: Components of MetS Outcomes Results: Weight Outcomes Results: Depression and Stress Outcomes
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended		Table 2 Supplemental Tables S4-S7
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory		Results: Components of MetS Outcomes; Weight Outcomes; Supplemental Information: Weight Outcomes
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms[6])		None reported
Discussion				
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses		Discussion: Limitations
Generalizability	21	Generalizability (external validity, applicability) of the trial findings	Describe key aspects of the setting which determined the trial results. Discuss possible differences in other settings where clinical tradi-tions, health service organization,	Discussion: Limitations

			staffing, or resources may vary from those of the trial.	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence		Discussion: Paragraphs 2, 4-6
Other information				
Registration	23	Registration number and name of trial registry		Abstract
Protocol	24	Where the full trial protocol can be accessed, if available		Methods; Proeschold-Bell RJ, Swift R, Moore HE, et al. Use of a randomized multiple baseline design: Rationale and design of the Spirited Life holistic health intervention study. *Contemp Clin Trials.* Jul 2013;35(2):138-152.
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders		Abstract; Acknowledgments

Supplement Table 3. CONSORT statement for abstracts[13]

Item	Standard Checklist item	Included
Authors	Contact details for the corresponding author	Y
Title	Identification of study as randomized	Y
Trial design	Description of the trial design (e.g. parallel, cluster, non-inferiority)	Y
Methods		
Participants	Eligibility criteria for participants and the settings where the data were collected	Y
Interventions	Interventions intended for each group	Y
Objective	Specific objective or hypothesis	Y
Outcome	Clearly defined primary outcome for this report	Y
Randomization	How participants were allocated to interventions	Y
Blinding (masking)	Whether or not participants, care givers, and those assessing the outcomes were blinded to group assignment	Y
Results		
Numbers randomized	Number of participants randomized to each group	Y
Recruitment	Trial status[1]	Y
Numbers analyzed	Number of participants analyzed in each group	Y
Outcome	For the primary outcome, a result for each group and the estimated effect size and its precision	Y
Harms	Important adverse events or side effects	NA
Conclusions	General interpretation of the results	Y
Trial registration	Registration number and name of trial register	Y
Funding	Source of funding	Y

[1] Relevant to Conference Abstracts

Supplement Table 4. Summary statistics of main health outcomes by time and randomized cohort (N=1,054)

	Immediate-Intervention Cohort (C1) (N=367)	One-year Waitlist Cohort (C2) (N=259)	Two-year Waitlist Cohort (C3) (N=428)
Time point (T)	**Prevalence: % (n/N)**	**Prevalence: % (n/N)**	**Prevalence: % (n/N)**
Metabolic syndrome			
T0, Baseline	49.5 (180/364)	49.4 (126/255)	51.7 (218/422)
T1, 12-month	40.5 (145/358)	49.8 (127/255)	49.4 (207/419)
T2, 18-month	38.7 (120/310)	N/A	51.5 (206/400)
T3, 24-month	42.9 (134/312)	42.7 (91/213)	49.6 (188/379)
T4, 30-month	N/A	44.6 (91/204)	N/A
T5, 36-month	N/A	46.1 (94/204)	45.3 (154/340)
T6, 42-month	N/A	N/A	47.1 (155/329)
T7, 48-month	N/A	N/A	45.1 (133/295)
Central obesity			
T0, Baseline	82.7 (301/364)	82.5 (212/257)	78.6 (331/421)
T1, 12-month	75.1 (269/358)	80.0 (204/255)	75.9 (318/419)
T2, 18-month	79.0 (245/310)	N/A	82.8 (331/400)
T3, 24-month	78.6 (246/313)	72.8 (155/213)	81.3 (309/380)
T4, 30-month	N/A	75.0 (153/204)	N/A
T5, 36-month	N/A	73.8 (152/206)	80.9 (275/340)
T6, 42-month	N/A	N/A	85.8 (283/330)
T7, 48-month	N/A	N/A	81.0 (243/300)
Elevated triglycerides			
T0, Baseline	48.6 (177/364)	50.0 (127/254)	54.2 (227/419)
T1, 12-month	38.9 (130/334)	47.3 (115/243)	48.7 (192/394)
T2, 18-month	35.6 (106/298)	N/A	45.1 (176/390)
T3, 24-month	42.2 (130/308)	38.7 (82/212)	44.7 (168/376)
T4, 30-month	N/A	38.6 (78/202)	N/A
T5, 36-month	N/A	49.0 (99/202)	47.3 (160/338)
T6, 42-month	N/A	N/A	41.8 (137/328)
T7, 48-month	N/A	N/A	44.0 (128/291)
Low high-density lipoprotein			
T0, Baseline	53.4 (194/363)	54.3 (138/254)	61.7 (258/418)
T1, 12-month	58.7 (210/358)	61.4 (156/254)	68.5 (287/419)
T2, 18-month	52.6 (163/310)	N/A	66.0 (264/400)
T3, 24-month	50.3 (157/312)	52.6 (112/213)	62.0 (235/379)
T4, 30-month	N/A	52.0 (106/204)	N/A
T5, 36-month	N/A	52.0 (106/204)	57.5 (195/339)
T6, 42-month	N/A	N/A	58.8 (193/328)
T7, 48-month	N/A	N/A	56.5 (166/294)
Hypertension			
T0, Baseline	52.9 (193/365)	50.4 (130/258)	51.8 (219/423)
T1, 12-month	42.5 (152/358)	52.2 (133/255)	53.2 (223/419)
T2, 18-month	42.1 (130/309)	N/A	49.1 (196/399)
T3, 24-month	42.8 (134/313)	43.7 (93/213)	52.5 (199/379)
T4, 30-month	N/A	43.6 (89/204)	N/A

T5, 36-month	N/A	51.0 (105/206)	46.2 (157/340)
T6, 42-month	N/A	N/A	45.5 (150/330)
T7, 48-month	N/A	N/A	48.2 (145/301)
Abnormal glucose regulation			
T0, Baseline	11.3 (41/362)	16.5 (41/249)	13.3 (55/413)
T1, 12-month	9.8 (35/358)	14.9 (38/255)	12.9 (54/419)
T2, 18-month	10.3 (32/310)	N/A	11.5 (46/400)
T3, 24-month	10.9 (34/313)	14.6 (31/213)	9.9 (37/375)
T4, 30-month	N/A	12.7 (26/204)	N/A
T5, 36-month	N/A	13.7 (28/204)	11.8 (40/340)
T6, 42-month	N/A	N/A	12.5 (41/329)
T7, 48-month	N/A	N/A	12.4 (37/299)
Depression (PHQ-8 ≥ 10)			
T0, Baseline	14.9 (54/363)	10.5 (27/258)	8.5 (36/423)
T1, 12-month	9.1 (31/339)	9.3 (22/236)	8.6 (36/417)
T2, 18-month	9.3 (30/322)	N/A	8.4 (34/407)
T3, 24-month	7.8 (23/295)	9.0 (19/210)	7.1 (28/392)
T4, 30-month	N/A	7.6 (15/198)	N/A
T5, 36-month	N/A	5.8 (12/206)	7.8 (27/345)
T6, 42-month	N/A	N/A	6.3 (20/315)
T7, 48-month	N/A	N/A	5.3 (16/303)
	Mean (SD), N	**Mean (SD), N**	**Mean (SD), N**
Perceived stress			
T0, Baseline	13.0 (6.2), 365	12.5 (6.2), 259	12.5 (6.1), 424
T1, 12-month	12.6 (6.0), 338	13.1 (5.9), 236	12.6 (6.0), 418
T2, 18-month	12.4 (6.1), 324	N/A	11.9 (5.7), 410
T3, 24-month	12.0 (6.1), 297	12.5 (6.4), 211	11.9 (6.3), 395
T4, 30-month	N/A	11.6 (6.7), 197	N/A
T5, 36-month	N/A	10.4 (6.5), 205	11.2 (6.7), 345

Notes:

Participants from whom data were collected for at least one follow-up wave were included for this analysis.

Perceived stress data were not collected after T5.

Data were not included if participant was pregnant or within 6 months postpartum at the time point of data collection.

For each cohort, the intervention period is shaded.

Supplement Table 5. Weight (KG) by time and randomized cohort

	Immediate-Intervention Cohort (C1)	One-year Waitlist Cohort (C2)	Two-year Waitlist Cohort (C3)
Time point (T)	**Mean (Standard deviation) [N]**		
All participants [N=1,054]			
	[N=367]	[N=259]	[N=428]
T0, Baseline	95.0 (24.0) [365]	93.0 (23.6) [258]	94.2 (23.4) [421]
T1, 12-month	92.2 (23.1) [358]	92.4 (22.6) [255]	93.9 (23.0) [419]
T2, 18-month	93.0 (24.5) [310]	N/A	93.7 (22.6) [400]
T3, 24-month	92.0 (22.6) [313]	87.6 (20.7) [213]	92.2 (22.0) [379]
T4, 30-month	N/A	87.4 (20.6) [204]	N/A
T5, 36-month	N/A	88.0 (21.6) [206]	90.8 (22.0) [339]
T6, 42-month	N/A	N/A	90.7 (21.9) [330]
T7, 48-month	N/A	N/A	90.5 (20.9) [300]
Participants who were obese (BMI≥30kg/m^2) at baseline [N=505]			
	[N=186]	[N =118]	[N=201]
T0, Baseline	111.8 (21.4) [186]	111.0 (21.8) [118]	110.6 (21.1) [201]
T1, 12-month	107.2 (21.6) [182]	109.8 (20.0) [115]	110.2 (20.8) [196]
T2, 18-month	108.6 (23.4) [159]	N/A	109.7 (20.7) [185]
T3, 24-month	106.7 (20.6) [160]	103.9 (17.9) [94]	107.7 (20.2) [173]
T4, 30-month	N/A	103.3 (18.0) [89]	N/A
T5, 36-month	N/A	104.9 (19.4) [89]	105.5 (21.0) [158]
T6, 42-month	N/A	N/A	106.0 (20.6) [150]
T7, 48-month	N/A	N/A	105.1 (19.5) [131]
Participants who were overweight (25 kg/m^2≤BMI<30 kg/m^2) at baseline [N=358]			
	(N=132)	(N=96)	(N=151)
T0, Baseline	83.6 (8.3) [122]	83.3 (9.1) [90]	84.8 (9.7) [146]
T1, 12-month	81.0 (9.8) [122]	83.4 (10.7) [89]	84.9 (10.1) [144]
T2, 18-month	80.7 (10.0) [104]	N/A	85.0 (10.0) [140]
T3, 24-month	81.0 (9.4) [105]	79.4 (10.2) [78]	84.5 (9.9) [132]
T4, 30-month	N/A	79.8 (11.2) [74]	N/A
T5, 36-month	N/A	79.9 (11.2) [74]	82.3 (10.9) [116]

T6, 42-month	N/A	N/A	82.2 (10.8) [115]
T7, 48-month	N/A	N/A	83.5 (11.1) [110]

Notes:

Participants from whom data were collected for at least one follow-up wave were included for this analysis.
Data were not included if participants were pregnant or within 6 months postpartum at the time of data collection.
For each cohort, the intervention period is shaded.

Supplement Table 6. Prevalences of three percent and five percent weight loss by time and randomized cohort (N=1,044)

	Immediate-Intervention Cohort (C1), N=365	One-year Waitlist Cohort (C2), N=258	Two-year Waitlist Cohort (C3), N=421
Time point (T)	**% (n/N)**		
Weight ≤ 97% of baseline weight			
T1, 12-month	46.4 (166/358)	20.5 (52/254)	16.7 (69/414)
T2, 18-month	43.0 (133/309)	N/A	20.8 (82/394)
T3, 24-month	47.3 (147/311)	52.6 (112/213)	30.8 (115/373)
T4, 30-month	N/A	53.4 (109/204)	N/A
T5, 36-month	N/A	49.8 (102/205)	47.6 (158/332)
T6, 42-month	N/A	N/A	43.0 (139/323)
T7, 48-month	N/A	N/A	37.2 (110/296)
Weight ≤ 95% of baseline weight			
T1, 12-month	33.0 (118/358)	10.6 (27/254)	9.9 (41/414)
T2, 18-month	32.7 (101/309)	N/A	12.7 (50/394)
T3, 24-month	34.1 (106/311)	44.1 (94/213)	19.8 (74/374)
T4, 30-month	N/A	40.7 (83/204)	N/A
T5, 36-month	N/A	39.0 (80/205)	33.7 (112/332)
T6, 42-month	N/A	N/A	30.0 (97/323)
T7, 48-month	N/A	N/A	27.4 (81/296)

Notes:

Participants from whom data were collected for at least one follow-up wave were included for this analysis.
Participants who missed baseline weight were excluded for this analysis.
Eight participants were pregnant or within 6 months postpartum at baseline, and therefore excluded for the analysis.
Data were not included if participant were pregnant or within 6 months postpartum at the time of data collection.
For each cohort, intervention period is shaded.

Supplement Table 7. Effectiveness of the Spirited Life intervention on weight by intervention duration (N=1,054)

	12-Month Intervention Effect	**18-Month Intervention Effect**	**24-Month Intervention Effect**
Weight (KG)	Coef (95% CI)*; p		
All participants	**-2.38 (-2.96, -1.81)***; p<0.001**	**-2.05 (-3.10, -1.01)***; p<0.001**	**-1.75 (-2.76, -0.74)***; p<0.001**
Participants obese at baseline	**-2.94 (-3.97, -1.91)***; p<0.001**	**-2.46 (-4.30, -0.62)**; p=0.009**	**-1.81 (-3.62, -0.01)*; p=0.048**
Participants overweight at baseline	**-2.49 (-3.26, -1.72)***; p<0.001**	**-2.42 (-3.56, -1.27)***; p<0.001**	**-2.63 (-3.81, -1.44)***; p<0.001**
Weight ≤ 97% of baseline weight	PR (95% CI)*; p		
All participants	**1.92 (1.67, 2.20)***; p<0.001**	**1.94 (1.56, 2.42)***; p<0.001**	**1.70 (1.43, 2.02)***; p<0.001**
Weight ≤ 95% of baseline weight	PR (95% CI)*; p		
All participants	**2.33 (1.95, 2.80)***; p<0.001**	**2.38 (1.79, 3.17)***; p<0.001**	**2.01 (1.60, 2.54)***; p<0.001**

Notes: Coefficients for each intervention level (12 months, 18 months, or 24 months in intervention vs no intervention) are estimated using ordinary least square linear regression modeling for weight, adjusting for time, district, and the baseline weight. Prevalence ratios for each intervention level are estimated using GEE Poisson regression modeling for binary outcomes (weight loss for 3% or more of baseline weight and weight loss for 5% or more of baseline weight), adjusting for time, district, and the baseline weight. Ninety-five percent confidence intervals are reported. Boldface indicates statistical significance (*p<0.05, **p<0.01, ***p<0.001).

References

1. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol.* 2009;62:464-475.
2. Koppenaal T, Linmans J, Knottnerus JA, Spigt M. Pragmatic vs. explanatory: an adaptation of the PRECIS tool helps to judge the applicability of systematic reviews for daily practice. *J Clin Epidemiol.* 2011;64(10):1095-1101.
3. Alberti KGMM, Zimmet P, Shaw J. Metabolic syndrome—a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med.* 2006;23(5):469-480.
4. The_International_Expert_Committee. International Expert Committee report on the role of the A1c assay in the diagnosis of diabetes. *Diabetes Care.* 2009;32(7):1-8.
5. Alberti K, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome a joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; American heart association; world heart federation; international atherosclerosis society; and international association for the study of obesity. *Circulation.* 2009;120(16):1640-1645.
6. Proeschold-Bell RJ, LeGrand SH. High rates of obesity and chronic disease among United Methodist clergy. *Obesity.* 2010;18(9):1867-1870.
7. Proeschold-Bell RJ, Miles A, Toth M, Adams C, Smith BW, Toole D. Using effort-reward imbalance theory to understand high rates of depression and anxiety among clergy. *The Journal of Primary Prevention.* 2013;34(6):439-453.
8. Zeger SL, Liang K-Y, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics.* 1988:1049-1060.
9. Hardin JW, Hilbe JM. *Generalized estimating equations.* 2 ed: Chapman and Hall/CRC; 2012.
10. Jolani S, van Buuren S. Doubly Robust Imputation of Incomplete Binary Longitudinal Data. *Journal of Biometrics & Biostatistics.* 2014;2014.
11. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *The Lancet.* 2008;371(9609):281-283.
12. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ.* 2008;337:a2390.
13. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med.* 2008;5(1):e20.