



**Comparison of Depression Identification
after Acute Coronary Syndrome:
Quality of Life and Cost Outcomes**
ClinicalTrials.gov: NCT01993017

STATISTICAL ANALYSIS PROTOCOL

LAST UPDATED: 1/1/2019

Table of Contents

1. SYNOPSIS OF THE STUDY	P.3
2. STUDY DESIGN AND OBJECTIVES	P.3
2.1 Study Design	
2.2 Objectives	
3. RANDOMIZATION AND BLINDING	P.3
4. DEFINITION OF TARGET POPULATION AND STUDY SAMPLES	P.4
4.1 Target Population	
4.2 Intent-to-Treat Sample	
4.3 Safety Analysis Sample	
5. DESIGN CONSIDERATION	P.6
5.1 Primary Statistical Hypothesis	
5.2 Interim and Final Primary Analysis	
5.3 Primary Efficacy Analysis	
5.4 Sample Size Determination	
5.5 Multiplicity	
5.6 Missing Data	
6. BASELINE CHARACTERISTICS	P.8
7. SAFETY/ TOLERABILITY	P.8
7.1 Primary (pre-specified) Safety Outcomes	
7.2 Mortality	
8. SECONDARY EFFICACY ANALYSES	P.8
9. REFERENCES	P.10

1. SYNOPSIS OF THE STUDY

The goal of this study is to determine the quality-adjusted life year benefits (primary outcome) and health care costs of following the AHA's advisory for depression screening and then referral for further diagnosis and treatment in post-ACS patients, if elevated depressive symptoms are found. To accomplish this aim, we will randomize patients from four different, geographically diverse health systems to three different groups: 1) to the AHA depression screen, notify, and treat if elevated depressive symptoms are found algorithm (Screen, Notify & Treat intervention group) or: 2) to receive no depression screening (No Screen; strong control group) or: 3) to be screened and a primary care provider notified of elevated depressive symptoms (Screen & Notify; minimally enhanced control group). Health-related quality of life, depressive symptoms, and costs will be obtained from all patients, so that the benefits and the costs of these three different depression screening strategies can be compared.

2. STUDY DESIGN AND OBJECTIVES

2.1 Study Design

Investigator initiated, multicenter, 3-group (1:1:1) randomized clinical trial.

2.2 Objectives

The overarching goal of this research is to conduct a state-of-the-art RCT that will rigorously evaluate the benefits and costs of AHA's depression advisory for modern post-ACS patients.

To examine in a randomized controlled trial the benefits and costs of the AHA's advisory for depression screen and treatment of post-ACS patients.

Hypothesis 1: Screen, notify, & treat intervention group will gain significantly more quality-adjusted life years (primary outcome) across 18 months when compared to No Depression screen control group, and also when compared to the Depression screen & notify control group.

Hypothesis 2: Those randomized to AHA's Depression screen & treat intervention group will have a favorable incremental cost-effectiveness ratio when compared to No Depression screen control and also when compared to the Depression screen & notify control group.

3. RANDOMIZATION AND BLINDING

Sequence generation

Participants will be randomly assigned to one of three groups: No Depression Screen, Depression Screen and Notify, or AHA Depression Screen, Notify and Treat. The randomization algorithm will be embedded in the web-based tracking system, using randomly assigned block sizes of 3, 6 and 9.

Concealment Mechanism

Participants will be randomized using the web-based tracking system. "User-roles" are assigned to study personnel, and the randomization tool is only available to unblinded coordinators at each site. Concealment will be ensured as the randomization algorithm will run in the backend, and only the randomization assignment will be visible to the unblinded coordinator after all necessary information about the participant has been entered.

Implementation

All eligible participants who give consent will be randomized by the unblinded coordinator after completion of all baseline assessments. The randomization tool is only available to designated unblinded coordinators (UC) at each site. The UC will enter the required information in the tracking system, after which the participant's group assignment will be immediately available.

If the participant is randomized to the No Depression Screen arm, the UC will inform the participant that no further assessments are required.

If the participant is randomized to the Depression Screen and Notify arm, the UC will administer the PHQ-8. If the PHQ score is ≥ 10 , the UC will notify the participant's primary care provider and/or cardiologist either through the tracking system, by email, or mail in accordance with their IRB requirements.

If the participant is randomized to the AHA Depression Screen, Notify, and Treat arm, the UC will administer the PHQ-8. If the PHQ score is ≥ 10 , the UC will inform the patient of the available treatment options. Depending on the participant's choice of treatment, the UC will facilitate the initial contact with the relevant treatment specialist(s), ideally within two weeks of randomization and will work closely with the treatment specialist(s) to monitor the participant's treatment progress. The participant's PCP and/or cardiologist will also be notified of the positive depression screen.

Blinding

The blinded coordinator (BC) will administer all study regular assessments at 6-, 12- and 18-months, and will not be allowed to know the participants' group allocation, and this is ensured by the "user-role" designation in the tracking system. In addition to the UC, and due to the nature of the study treatment, participants, site PIs, and other personnel not designated as BC, cannot be blinded to the group allocation, but are encouraged not to disclose the allocation either at or when assisting the BC in scheduling follow up assessments. After each study visit, the BC will complete a 1-item questionnaire asking if the participant disclosed if they were in treatment. This information will be tracked, however no intervention will occur if the BC becomes unblinded.

4. DEFINITION OF TARGET POPULATION AND STUDY SAMPLES

4.1 Target Population

This study aimed to enroll survivors of acute coronary syndromes without a prior history of depression who had experienced an ACS within the past 2 to 12 months and would be eligible for depression screening.

Detailed eligibility criteria are provided below.

Inclusion Criteria

Criteria	EMR Verification and ICD-9 Codes
English or Spanish*-speaking participants *Spanish-speaking participants eligible at Columbia-site, only	Primary language designation in EMR (if available) Participant attestation

Documented ACS within the past 2-12 months	Evidence of one or more of the following within the past 2-12 months: (1) Diagnosis of acute myocardial infarction (410) during an inpatient hospitalization (2) Diagnosis of unstable angina (411) during an inpatient hospitalization with a history of coronary artery disease (414) 410.00-410.92 Acute Myocardial Infarction 411.00-411.89 Other acute and sub-acute forms of ischemic heart disease 414.00-414.9 Other forms ischemic heart disease
Over the age of 21 years	DOB
Has access to a phone and/or computer	Participant confirmation of phone number listed in EMR

ICD 9 discharge codes of 410 (acute myocardial infarction) through EMR searches have excellent positive predictive value when clinical data are abstracted and checked by two coders blinded to discharge code. Thus, we will use this discharge code for ACS eligibility. We will furthermore select potential participants with ICD 9 code hospital discharge codes of 411 (unstable angina), who also have established coronary artery disease (ICD 9 code of 414) to ensure that participants meet the definition of an ACS, as described in ACS case definitions by numerous cardiology societies. This approach – having broad eligibility – provides for a high degree of generalizability.

Medical Exclusion Criteria

Terminal illness defined as, but not limited to:
NYHA class IV, ACC class D CHF requiring inotropes or mechanical assist devices or critical aortic stenosis without plan for correction
End-stage COPD/emphysema
Advanced cirrhosis with encephalopathy, varices, severe ascites
Severe rheumatologic diseases requiring frequent hospitalizations, and multiple cytotoxic agents and/or disease modifying drugs
Metastatic pancreatic, esophageal, colorectal or stomach cancer
Metastatic sarcoma, ovarian, melanoma or renal cell cancer
Metastatic breast cancer with multiple recurrences despite treatment
Advanced CNS malignancies
Recurrent hematologic malignancies with multiple recurrences despite treatment
Persistent AIDS, untreated or treated
Currently pregnant

Participant Reported Screening Exclusions

Potential participants will complete a brief screening questionnaire to confirm that they have none of the above conditions, that they speak English and that they are interested in being enrolled in the study

4.2 Intent-to-Treat Sample

Analyses will be conducted on an intent-to-treat sample. No analyses will be conducted on a per-protocol or safety analysis sample.

4.3. Safety Analysis Sample

Safety analyses will be conducted using the same sample as the primary efficacy sample, and an intent-to-treat analysis plan will be used.

5. DESIGN CONSIDERATION

5.1 Primary Statistical Hypothesis

Those randomized to AHA's Depression Screen, Notify & Treat intervention group will gain significantly more quality-adjusted life years across 18 months when compared to No Depression Screen control group, and also when compared to the Depression Screen & Notify control group.

5.2 Interim and Final Primary Analysis

Given the nature of a screening trial, no *a priori* stopping rules were planned. Interim analyses were planned for purposes of monitoring the study for safety concerns, only.

5.3 Primary Efficacy Analysis

Change in QALYs from baseline through 18 months post-randomization will serve as the primary outcome for this trial. QALYs were chosen as the primary outcome to facilitate comparisons of the effect of depression screening with other preventive interventions as well as to facilitate cost-effectiveness analyses and policy decisions. As QALYs do not directly assess depression, this outcome measure also minimized possible patient reporting bias as a result of lack of participant blinding or masking to condition. The goal of the primary analysis is to identify whether there is difference in the change in QALYs among the three groups. QALYs describe the duration of illness per years of survival, adjusted for quality of life experienced during that survival. One year in perfect health is equivalent to 1 QALY. All patients will complete a standardized measure of quality of life using the Short Form-12 Health Survey, Version 2™ (SF-12) at baseline, and again at 6-months, 12-months, and 18-months.¹ QALYs will be estimated from the SF-12 using the Short Form 6 Duration (SF6D) which converts data from 7 items in the SF-12 assessing 6 domains (physical functioning, role limitations, social functioning, pain, mental health and vitality) to QALYs.² Study patients who die during the study period will be assigned a utility score of 0 for assessments after the date of death. QALYs over 18 months will be calculated as the area-under-curve by interpolating linearly the scores at the four assessments (baseline, 6-month, 12-month, 18-month). There were no data obtained or available for these measures between assessments. Change in QALYs will then be obtained by subtracting the QALYs that would have occurred if there was no change in the baseline utility score from the actual QALYs that were measured across 18 months. The analysis will follow the principle of intention-to-treat and will be conducted on each imputed dataset with the point

estimate deriving from the average of 5 datasets and the pooled variance calculated using Rubin's formula.

To determine the significance of differences, we will perform a two-step gate-keeping test procedure. This will involve first performing an F-test using ANOVA, and then proceeding to do all three pairwise comparisons using two-sided t-test at 5% nominal significance *only if* the F-test has a p-value less than 0.05. With three randomization groups and three pairwise comparisons, this two-step procedure can be shown to preserve the familywise error rate in the strong sense, that is, a false positive comparison will occur at most with 5% probability under all possible scenarios.^{3,4} This method is also generally more powerful than Bonferroni's adjustments.

5.4 Sample Size Determination

The sample size of the trial was determined based on an assumed standard deviation for QALYs of 0.17.⁵ Additionally, based on a general health-related quality of life outcome in a prior study of management of depression for patients with cancer, we assumed a net improvement in QALYs of 0.155 over 18 months of follow-up for depressed individuals who receive depression treatment in the Screen, Notify and Treat group.⁶ We assumed a 0.055 gain in QALYs for all patients not directly linked to depression treatment (i.e., those in the No Screen group, the Screen & Notify group and those without elevated depressive symptoms in the Screen, Notify, and Treat group). An important consideration for this trial is that only 20% of patients randomized to the Screen, Notify and Treat group were expected to meet criteria for elevated depressive symptoms and thus, to receive depression treatment. Therefore, assuming an increase in QALYs of 0.21 (0.055 background improvement + net improvement of 0.155) over the 18-month follow-up period for the 20% of patients diagnosed and treated for depression in the Screen, Notify, and Treat group and a 0.055 improvement in QALYs for the 80% of patients in this randomization group without depression, an overall gain in QALYs of 0.086 over the 18 month follow-up period was anticipated in this randomization group ($0.21 * 0.2 + 0.055 * 0.8 = 0.086$). Thus, we anticipated a difference in QALYs of 0.031 (0.086 change in the Screen, Notify, and Treat group minus 0.055 in the No Screen group or Screen and Notify Group), leading to an expected effect size of 0.18 ($= 0.031/0.17$). With this effect size, we determined the sample size per group to be 475, which would yield 80% power for a two-sided t-test at 5% level. We chose to determine sample size based on a pairwise comparison at 80% power in the two-step procedure (described above), as a conservative approach relative to powering based on the F-test. Specifically, under the scenario where one group has higher QALY than the other two by an effect size 0.18, the F-test will yield 84% power. Adding in 5% loss to follow-up, we selected an overall sample size of n=500 in each randomization group for an overall sample size of n=1500.

5.5 Multiplicity

Our use of the 2-step gatekeeping function will enable us to preserve the familywise error rate in the strong sense, that is, a false positive comparison will occur at most with 5% probability under all possible scenarios.^{3,4} This method will enable us to consider multiple group comparisons without using Bonferroni's adjustments.

5.6 Missing Data

CESD-10 questionnaire with missing data will be prorated if that participant answered more than 7 items. The missingness of data will then be assessed using Little's test. If data are found to be missing at random, we will perform multivariate imputations to generate 5 imputed datasets by basing on covariates which were predictive of missing pattern such as input from prior visit and site variable; random sampling was used to impute the missing values at baseline. Missing data will be imputed sequentially, starting with the baseline visit, then the 6-month visit, followed by the 12-month and 18-month visits.

In sensitivity analyses, missing data will be handled by carrying the last observed value forward and by using best case-worst case scenario – in best case, we assumed that all missing had perfect health. In worst case scenario, we assumed all missing data had worst health status (QOL = 0).

6. BASELINE CHARACTERISTICS

Baseline characteristics will be examined as means (standard deviation) or percentages by randomization assignment to assess for a balanced allocation.

7. SAFETY/ TOLERABILITY

7.1 Primary (pre-specified) Safety Outcomes

Harms attributable to use of antidepressant medications (i.e., appetite problems, sleep problems, gastrointestinal upset, and bleeding) will be assessed through patient interview. Group differences in the prevalence of these potential adverse effects will be compared using chi-squared tests.

7.2 Mortality

Mortality will be assessed by surveying patient surrogates and through review of the electronic medical record. Group differences in mortality will be compared using chi-squared tests.

8. SECONDARY EFFICACY ANALYSES

Key secondary outcomes included depression-free days and health care costs.

Depression-Free Days

The prespecified secondary outcome will be cumulative depression-free days based on the 10-item Center for Epidemiologic Studies Depression (CESD-10) scale, a non-diagnostic, epidemiologic, reliable and valid measurement for depressive symptoms, measured at baseline, 6-months, 12-months, and 18-months in all 3 arms.⁷ *Depression-free days* will be calculated by using linear interpolation to estimate daily depression severity at each of the four assessment time points for the 3 study arms. Depression-free days is a valid and easily interpretable measure for estimating depression treatment outcomes when multiple measures of depressive symptoms occur over time. This measure is also amenable to cost-effectiveness analyses. In other trials, Depression-free days have been calculated using intervals as long as 6 months.⁸ While a shorter interval (e.g., 3 months) can provide a more precise assessment of cumulative depressive symptoms over time, a 6-month interval was selected so as to avoid frequent

communication between the study team and study participants. More frequent assessments could lead to increased behavioral support and could cloud the interpretation of the No Depression Screen group which was not intended to receive any behavioral interventions.

An epidemiologic instrument was chosen over a clinical measure of depressive symptoms as the use of a clinical measure would have mandated referral for depression treatment in those who screened positive in the No Screen group. The CESD-10 is a short version of the original 20-item scale. The scores range from 0 to 30. A score 4 or greater has been found to be the optimal cutpoint for a positive depression screen with a sensitivity and specificity of 97% and 84%, respectively, compared to a psychiatric interview in a sample of older adults.⁹ Based on a review of the literature, we inferred that a cutpoint of 10 on the 10-item CESD would represent clinically significant depressive symptoms.¹⁰

The following rule will be used to convert CESD score to depression day

CESD < 4 -> 0 depression day
CESD = 4 -> 1/7 depression day
CESD = 5 -> 2/7 depression day
CESD = 6 -> 3/7 depression day
CESD = 7 -> 4/7 depression day
CESD = 8 -> 5/7 depression day
CESD = 9 -> 6/7 depression day
CESD >= 10 -> 1 depression day

Health Care Costs and Lost Productivity

The other prespecified secondary outcome will be Health Care Costs and Lost Productivity.¹ At baseline, 6-months, 12-months, and 18-months, patients will report measures of economic productivity, including employment status, occupation, hours spent at work, and time lost from work for health-related reasons.¹¹ At 6-months, 12-months, and 18-months, patients will also report healthcare utilization since their last intake assessment, including emergency department (ED) visits, hospitalizations (location, admission and discharge dates), psychiatric medication use, name and dose, ambulatory care visits with mental health specialists, cardiologists, as well as PCPs and finally hospitalizations for cardiovascular events. Patient self-reports of healthcare utilization will be supplemented by review of the EMR and claims systems to collect data on healthcare utilization during the 18-month trial period. Average Medicare reimbursement rates according to diagnosis-related groups will be applied to inpatient visits to estimate hospitalization costs, and the Medicare physician fee schedule will be applied to outpatient and ED resource use according to current procedural terminology codes. Costs of study depression treatment will also be incorporated into estimates of healthcare utilization costs for those assigned to the Screen, Notify, and Treat group who agree to depression treatment by study personnel. To estimate economic costs from a societal perspective, changes in productivity and time spent traveling to appointments will also be accounted for. Costs will be standardized across years using the U.S. Consumer Price Index and presented in U.S. dollars.¹²

¹Cost of health care utilization was erroneously listed as a co-primary outcome on the initial clinicaltrials.gov entry. This error was corrected in clinicaltrials.gov on October 25, 2017, prior to completion of data collection or any interim data analyses. As can be seen in the planned statistical analyses, the sample size for the study was driven entirely by the change in QALYs outcome. Cost-effectiveness analyses, in contrast, were never intended to be listed as a co-primary outcome as they were not guided by economic hypotheses involving statistical tests and sample size calculations for incremental cost-effectiveness ratios were never attempted.

9. REFERENCES

1. Ware J, Kosinski M, Keller S. *How to Score the SF-12 Physical and Mental Health Summaries: A User's Manual*. Boston, MA: The Health Institute, New England Medical Centre; 1995.
2. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851-859.
3. Cheung YK. Sequential Implementation of Stepwise Procedures for Identifying the Maximum Tolerated Dose. *Journal of the American Statistical Association*. 2007;102(480):1448-1461.
4. Hochberg Y, Tamhane AC. *Multiple comparison procedures*. John Wiley & Sons, Inc.; 1987.
5. Davis JC, Marra CA, Najafzadeh M, Liu-Ambrose T. The independent contribution of executive functions to health related quality of life in older women. *BMC Geriatrics*. 2010;10(1):16.
6. Strong V, Waters R, Hibberd C, et al. Management of depression for people with cancer (SMaRT oncology 1): a randomised trial. *The Lancet*. 2008;372(9632):40-48.
7. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*. 1977;1(3):385-401.
8. Vannoy SD, Areal P, Unützer J. Advantages of Using Estimated Depression-Free Days for Evaluating Treatment Efficacy. *Psychiatric services (Washington, DC)*. 2010;61(2):160-163.
9. Irwin M, Artin KH, Oxman MN. Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Arch Intern Med*. 1999;159(15):1701-1704.
10. Björgvinsson T, Kertz SJ, Bigda-Peyton JS, McCoy KL, Aderka IM. Psychometric properties of the CES-D-10 in a psychiatric sample. *Assessment*. 2013;20(4):429-436.
11. Prasad M, Wahlqvist P, Shikhar R, Shih YC. A review of self-report instruments measuring health-related work productivity: a patient-reported outcomes perspective. *Pharmacoeconomics*. 2004;22(4):225-244.
12. Unutzer J HH, Schoenbau M., Druss, B.. The collaborative care model: An approach for integrated physical and mental health care in meidcad health homes. *Health Home Information Resource Center Brief* May 2013. <http://www.medicaid.gov/State-Resource-Center/Medicaid-State-Technical-Assistance/Health-Homes-Technical-Assistance/Downloads/HH-IRC-Collaborative-5-13.pdf> Accessed April 2015. .