Published in final edited form as:
Contemp Clin Trials. 2013 January ; 34(1): 10–20. doi:10.1016/j.cct.2012.08.011.

Comprehensive Assessment of Long-Term Effects of Reducing Intake of Energy Phase 2 (CALERIE Phase 2) Screening and Recruitment: Methods and Results

TM Stewart[1], **M Bhapkar**[2], **S Das**[3], **K Galan**[2], **CK Martin**[1], **L McAdams**[2], **C Pieper**[2], **L Redman**[1], **S Roberts**[3], **RI Stein**, **J Rochon**[2], and **DA Williamson**[1] **for the CALERIE Study Group***

[1]Pennington Biomedical Research Center, Baton Rouge, LA

[2]Duke University Medical Center, Durham, NC

[3]Tufts University, Boston, MA

[4]Washington University, St. Louis, MO

Abstract

The Comprehensive Assessment of the Long-term Effects of Reducing Intake of Energy Phase 2 (CALERIE) study is a systematic investigation of sustained 25% calorie restriction (CR) in non-obese humans. CALERIE is a multicenter (3 clinical sites, one coordinating center), parallel group, randomized controlled trial. Participants were recruited, screened, and randomized to the CR or control group with a 2:1 allocation. Inclusion criteria included ages 21–50 years for men and 21–47 years for women, and a body mass index (BMI) of $22.0 \leq \text{BMI} < 28.0$ kg/m^2. Exclusion criteria included abnormal laboratory markers, significant medical conditions, psychiatric/behavioral problems, and an inability to adhere to the rigors of the evaluation/intervention schedule. A multi-stage screening process (telephone screen and 3 inclinic visits) was applied to identify eligible participants. Recruitment was effective and enrollment targets were met on time. 10,856 individuals contacted the clinical sites, of whom 9,787 (90%) failed one or more eligibility criteria. Of the 1,069 volunteers who started the inclinic screening, 831 (78%) were either ineligible or dropped. 238 volunteers were enrolled (i.e., initiated the baseline evaluations), 220 were randomized, and 218 started the assigned intervention (2% from the first screening step). This study offered lessons for future multi-center trials engaging non-disease populations. Recruitment strategies must be tailored to specific sites. A multi-disciplinary screening process should be applied to address medical, physical, and psychological/behavioral suitability of participants. Finally, a multi-step screening process with simple criteria first, followed by more elaborate procedures has the potential to reduce the use of study resources.

Corresponding author information: Tiffany M. Stewart, Ph.D,. Director, Behavior Technology Laboratory: Eating Disorders & Obesity, The Pennington Biomedical Research Center, 6400 Perkins Rd., Baton Rouge, LA 70808, tiffany.stewart@pbrc.edu, Phone: 225-763-2554.

*A full listing of all CALERIE investigators and staff may be found elsewhere (1).

Keywords

Aging; Caloric restriction; randomized controlled trial; consort information; participant recruitment; participant screening

Introduction

The extent to which calorie restriction (CR) may have benefits on primary aging (age associated deficits in function) or secondary aging (natural history of common age-related morbidities) in humans is not known. Moreover, the extent to which sustained, human CR might prevent or delay the effects of primary aging but might be accompanied by unacceptable side-effects has not been investigated. Thus, the Comprehensive Assessment of the Long-term Effects of Reducing Intake of Energy (CALERIE) research program was designed to systematically investigate sustained calorie restriction (CR) in humans on markers of the aging process including resting metabolic rate, core temperature, oxyradical formation, cardiovascular risk markers, insulin sensitivity and secretion, immune function, neuroendocrine function, cognitive function, and quality of life. Phase 1 of the CALERIE program involved three pilot studies (1–3). The present study, CALERIE Phase 2, was designed to investigate the impact of CR in nonobese humans and was conducted at three clinical sites: Pennington Biomedical Research Center (PBRC), Baton Rouge, LA; Tufts University (Tufts), Boston, MA; and Washington University School of Medicine, St. Louis, MO; and was coordinated by Duke Clinical Research Institute, Durham, NC. A description of the CALERIE Phase 2 study design and methods has been reported (4).

Experience from the CALERIE Phase 1 studies guided the recruitment and screening procedures utilized in the CALERIE Phase 2 study to ensure that participants were qualified and could persevere to the rigorous CR intervention throughout the entire two year study (4). The CALERIE Phase 2 study aimed to recruit and randomize 225 participants, with approximately equal distribution across the three participating sites.

Due to the lack of research on human CR, little is known about recruitment and screening procedures for this type of intensive study. Further, little is known about individuals who are interested in participating in studies on CR. Thus, the primary aims of the present paper are to: 1) Describe the recruitment procedures utilized in the CALERIE Phase 2 study, 2) Describe the screening procedures in the study, 3) Describe the reasons for exclusion of participants from the study across the three sites, 4) Describe the results of the CALERIE Phase 2 recruitment and screening process, and 5) Describe differences in participants who were excluded/dropped out compared to those who were randomized in the study (external validity of the CALERIE Phase 2 cohort).

Methods

Eligibility criteria

Healthy individuals from both genders and all races were eligible to participate. Men were eligible to participate if they were between 21 and 50 years of age (inclusive); women were eligible if between 21 and 47 years of age (inclusive); and, $22.0 \leq \text{BMI} < 28.0$ kg/m^2. The trial attempted to be as inclusive as possible. The different age requirements for men and women were aimed at ruling out confounding due to the peri-menopausal effects in women (4). Of note was the CALERIE Phase 2 study was not a typical weight loss study in overweight individuals. Rather, Phase 2 was designed to study the physiological effects of caloric restriction in normal weight and slightly overweight individuals. The upper BMI

limit was selected to exclude obese individuals, and the lower BMI limit was selected primarily for safety reasons (4).

Volunteers were ineligible if there were significant medical conditions (e.g., history or clinical manifestation of cardiovascular disease, diabetes, cholelithiasis or cancer); abnormal laboratory markers (e.g., elevated potassium levels; hemoglobin, hematocrit, or iron below their respective lower limits of normal; LDL-cholesterol level ≥ 190 mg/dl; serum ALT > 1.5 × the upper limit of normal [ULN]); psychiatric or behavioral problems (e.g., history or clinical manifestation of any eating disorders, history of drug and alcohol abuse, pharmacologic treatment for a psychiatric disorder, Beck Depression Inventory (BDI) score ≥ 20); concomitant medications (e.g., short-term treatment with steroids within six months prior to randomization, regular use of other medications except contraceptives). Breast-feeding or pregnant women (or those intending to become pregnant before the scheduled end of the study), and individuals engaged in a regular program of vigorous physical fitness, were not eligible. Volunteers were ineligible if they were unwilling or expressed an inability to adhere to the rigors of the CR intervention or the evaluation schedule over the entire two-year period. Specific details regarding eligibility criteria are published elsewhere (4).

Recruitment procedures

Participants were recruited at the three CALERIE clinical sites (PBRC, Tufts, and Washington University). Recruitment was continuous, and included health promotion events, media advertising, direct mail, databases, and referral sources. A brochure was created summarizing the goals of CALERIE, study procedures, and measures to ensure participant safety. The recruitment materials for the study made it clear that the study was primarily a study of the potential beneficial effects of CR on health and biomarkers of aging in humans and not strictly a study of weight loss for overweight individuals. Based on the brochure and materials, a 12-minute video was created summarizing the purpose of the study, procedures, and safety issues. The video was made available on CALERIE's public website (http://calerie.dcri.duke.edu), and was presented to volunteers during their first in-clinic visit.

Recruiting was conducted independently at each site using approaches which have proven successful in their respective locations. Specifically, this included methods successfully applied in the CALERIE Phase 1 studies (1). In addition, the success of the different strategies was reviewed in study-wide teleconference calls to ensure that each site benefited from the collective experience.

PBRC—Recruitment activities at PBRC were coordinated by the Clinical Trial Recruitment Services Core (CTRSC) which included a call center staffed by six full time recruiters to manage incoming calls for all studies at PBRC. The CTRSC employed a mixed approach by utilizing established methods from the CALERIE Phase 1 study (1), such as direct mail and newspaper, while continuing to develop and deploy innovative advertisement media including a listserv application and online advertisements.

Information on age and BMI for over 60,000 people interested in research studies was available in the PBRC database. This database was queried and potential participants identified. Mail lists were purchased from a third party for zip codes in the surrounding areas of the research center. Letters were mailed in batches to purchased zip code mailing lists in the Baton Rouge area. Post cards were sent out in batches of a 1000 every 4 days to achieve a manageable load of incoming calls. Additionally, a center-wide email was sent to Pennington Center employees and information about the CALERIE study was printed on Louisiana State University (LSU) employee check stubs. Also, during the study, a "satellite"

recruiting and clinical trial site was opened, in partnership with the Baton Rouge General Hospital (BRGH), in downtown Baton Rouge.

Further, an advertisement tracking system provided the metrics needed to measure efficacy of advertisement. At PBRC, the advertisement tracking system consisted of the following. Phone screen numbers were logged and reviewed weekly. The total number of screens, including phone and in person were reviewed weekly and compared to the overall recruitment goals. Next, the percentage of screeners eligible to schedule Screening visit 1 (SV1) were reviewed. The goal was a high number of phone calls and high number of individuals eligible for SV1. Finally, ad content was changed when call volume slowed down or when there was a large number of individuals not eligible for SV1. Similar observations were made at Tufts and Washington University, e.g. the yield from various sources of advertising was used to inform and improve recruitment strategies.

Finally, PBRC included a comprehensive evaluation of the health literacy levels of PBRC advertisements. In response to this evaluation, all CALERIE advertisements were reviewed and adapted to match the appropriate literacy level (4^{th} grade reading level). In its first 9 months of use, the newly worded advertisements yielded a considerable increase in the number of phone screens.

Tufts—Recruitment efforts were coordinated by the investigators with staff support from the recruiting department including three designated part-time recruiters who were extensively trained in the details of the CALERIE protocol so they could effectively answer incoming calls and e-mails about the study. Methods of recruiting at Tufts included flyers, advertising in local newspapers and CraigsList website, contacting former volunteers for research studies at Tufts, and the Tufts website.

The Tufts Human Nutrition Research Center on Aging (HNRCA) recruiting office maintains a database of over 20,000 potential volunteers from which men and women meeting the age criteria could be identified. Eligible volunteers were sent a letter signed by the principal investigator explaining the study, and postage prepaid return envelopes were enclosed to facilitate a response. Advertising in a variety of media was conducted. This included advertisements in regional and community newspapers and health related articles written by the CALERIE staff periodically accompanied the newspaper section where the study was advertised. Advertisements were also placed on local radio and TV. These advertisements were strategically staggered using a monthly schedule to cover both weekdays and weekend days, and at different times during the day in order to reach a diverse audience. The placement of these advertisements were coordinated with times and programs geared towards particular demographic groups. The central website address for the study was included on all printed materials and flyers. Using the Tufts intranet email, announcements describing the study were sent periodically to all Tufts affiliated employees and students, including those at Tufts Medical Center and all of the professional schools. Posted flyers were also distributed at grocery stores, community gymnasiums, health clubs, corporations, churches, and other high traffic gathering areas to improve study visibility in Boston and the surrounding areas. Direct mailings were made to specific zip codes in the Boston area. A variety of local activities such as local community network systems, and all health fairs and community outreach efforts conducted by the HNRCA were used to reach the target population.

Washington University—Recruitment activities were performed by the Study Coordinator and Study Manager. The recruitment approach included mass e-mails sent to employees of Washington University and two affiliated hospitals, Barnes-Jewish Hospital and St. Louis Children's Hospital, as well as a series of news releases, radio, and TV

interviews with study investigators. The news releases and interviews with reporters resulted in publication of articles in St. Louis area newspapers and magazines. These articles highlighted the long-term effects of CR on health and longevity, described the CALERIE Phase 2 study, and included a phone number for individuals interested in the study to call.

Other volunteers/participants were recruited at health fairs, by means of posters placed at various sites in the medical center, by referral by current study participants, and mass mailings. Mailing labels for addresses in the St. Louis area were purchased and letters describing the study and inviting participation were mailed to individual zip codes surrounding Washington University. Batches of 50 to 100 letters were sent out, enabling the CALERIE study team to respond to the phone and e-mail inquiries in a timely manner. Local businesses, churches, colleges, and grocery stores were targeted with recruitment flyers. Finally, the “Volunteer for Health” program (operated by the Washington University Center for Clinical Studies and utilized for various studies at the medical school) was used as a recruitment mechanism. The program provided information to the public about ongoing research through display cases, flyers, and a website. In addition, this resource provided a large database that was searched for potential volunteers who met the study inclusion criteria and appropriate individuals were contacted.

Screening Procedures

The CALERIE Phase 2 study implemented a rigorous screening process to carefully identify potential participants. The goal was to screen out individuals who were either not eligible or unsuitable for the intervention or were unlikely to adhere to the rigors of the intervention and/or evaluation schedule. The details of the activities for each of the screening visits are outlined in Table 1. Volunteers could be deemed ineligible and/or choose to withdraw from screening at any time during the process.

In some cases, due to the need to clarify further information that may be a barrier to participation, a follow-up screening visit was added. At the end of the screening process, results from the assessments were summarized on screening checklists and these documents were used to provide an overall eligibility assessment for each candidate. The suitability of each candidate for CALERIE was then assessed by a multidisciplinary team of behavioral experts, dietitians, and clinical staff. If approved, the participant was invited to enroll in the study and begin the baseline evaluations. The study manager at each site informed candidates of their acceptance into the study and prepared for enrollment of the participant in baseline testing of the study. Ineligible candidates received a letter explaining the reasons for their ineligibility.

Assessment Measures

Assessment measures utilized in the CALERIE screening procedures are outlined below.

7-day Stanford Physical Activity Record (PAR; 5)—On the PAR, the participant reports the amount of time spent sleeping and participating in moderate, hard or very hard physical activities. The PAR was used to assess the activity level of each candidate to screen out those who are very physically active or are in training. The following criteria was generally used to establish exercise patterns intense enough to warrant exclusion: Individuals who were engaged in a regular program of physical fitness involving some kind of heavy physical activity (e.g., jogging, running or riding fast on a bicycle for 30 minutes or more) five or more times per week over the past year

Eating Inventory (EI; 6)—The EI assesses dietary restraint, disinhibition, and perceived hunger. The questionnaire produces three scores for the three subscales: dietary restraint,

disinhibition, and perceived hunger. The EI was used to measure these three variables in the study and has been found to be valid and reliable (6). These three variables were used to assess a pattern of eating that was queried further in the Barriers to Participation Interview described below.

Multiaxial Assessment of Eating Disorder Symptoms (MAEDS; 7)—The MAEDS is a 56 item self-report inventory that measures the six symptom domains related to eating disorders on separate subscales: binge eating, restrictive eating, purgative behavior, fear of fatness, avoidance of forbidden foods, and depression. The MAEDS has been shown to have adequate validity and reliability (7). The MAEDS was used as an efficient assessment of eating disorder symptoms. Subscale scores are norm-referenced. Participants who scored two standard deviations above the mean (t-score=70 or above) on any MAEDS subscale were given the Interview for the Diagnosis of Eating Disorders (IDED-IV) to detect the presence of eating disorder symptoms (8).

Structured Clinical Interview for Diagnosis of DSM-IV Personality Disorders Questionnaire (SCID-II; 9)—The SCID-II Personality Questionnaire is a screening tool to shorten the time it takes the clinician to administer the SCID-II Interview (9). After the subject fills out the questionnaire, the clinician needs only to inquire about the items screened positive on the questionnaire. This measure was used to screen for personality disorder symptoms. If a participant indicated personality disorder symptoms (indicated by achieving particular scores on each symptom as designated by the questionnaire), the full interview was given to assess symptoms further and determine if personality disorder symptoms posed a threat to a participant's adherence success in the study.

Beck Depression Inventory (BDI,10)—The BDI-II is an assessment of current depressive symptomology composed of 21 groups of statements. The participant reads each group and chooses the one statement from each group that most accurately reflects how they have been feeling for the last two weeks. The BDI has been found to have adequate validity and reliability (10).The BDI-II was used to assess current depression symptoms. Individuals who scored ≥ 20 at screening or baseline were excluded from the study. Further, if individuals self-reported clinically significant symptoms of depression on other measures, e.g. MAEDS and the Barriers to Participation Interview, then it was examined with regard to the impact of mood on the ability to adhere to study requirements.

Barriers to Participation Interview—The Barriers to Participation Interview is a standardized interview that was used to ascertain participants' willingness to commit to the study, support from household members to participate in the study, motivation and challenges facing the participant and other similar study specific issues of importance. Issues that posed a significant threat to study adherence that were exposed in the Barriers Interview were further queried and decisions were made about eligibility. Various issues included but were not limited to: 1) any history of pharmacologic treatment for a psychiatric disorder within one year prior to the randomization date or a history of more than one episode of a pharmacologic treatment for a psychiatric disorder within lifetime, 2)history of drug or alcohol abuse (up to 14 drinks a week are allowed) within the past two years, 3) individuals who practice a vegan dietary lifestyle , 4) unwilling to be assigned at random to the CR or control intervention, 5) unwilling or unable to adhere to the rigors of the CR intervention over the entire two-year intervention period, 6)individuals who unable or unwilling to discontinue dietary supplements or adhere to the alcohol consumption restrictions during the study, and 7) unwilling or unable to adhere to the rigors of the data collection and clinical evaluation schedule over the entire two-year period follow-up period

Body Morph Assessment 2.0 (11)—The BMA 2.0 measures estimates of perceived current body size, ideal body size, body size dissatisfaction (perceived current body size-ideal body size). The BMA 2.0 measures very small increments of changes in body size estimation. There are one hundred total increments from the extremely thin endpoint on the measure to the obese endpoint. It is computer-based and self-administered. The reliability and the validity of the BMA 2.0 has been supported (11). This measure served as a measurement for body dissatisfaction and body size over/under estimation in the present study. If a participant scored two standard deviations or more above the mean on the current body size scale of the BMA (2 standard deviations is equal to a *t*-score of 70, these people view their body as larger than it actually is), and/or scored two standard deviations or more below the mean (*t*-score < 30) on the Ideal Body Size scale (these people desire an ideal body size that is excessively small), an IDED interview was administered to query further body image and/or eating disorder symptoms.

The Interview for the Diagnosis of Eating Disorders (IDED-IV;8)—The IDED-IV is a semi-structured clinical interview that is used for the assessment for the presence or absence of eating disorder symptoms according to the DSM-IV. The IDED-IV has been found to have adequate internal consistency, interrater reliability, concurrent validity and discriminant validity for diagnosing eating disorders using DSM-IV criteria (8). Participants who were given the IDED-IV (as triggered by other measures, e.g. MAEDS, BMA, or Barriers to Participation Interview) that presented with a history or clinical manifestation of any clinical or subclinical eating disorder were excluded.

Statistical Analysis

Recruitment, Screening and Enrollment—This paper utilizes descriptive statistics to summarize the different steps in the screening and exclusion/inclusion process.

Reasons for Exclusion/Dropout—Descriptive statistics were used to summarize the reasons for exclusion from the study via different subgroups. For the exclusions, volunteers could have been ineligible on several criteria simultaneously. In these cases, the exclusion reason observed first was noted.

External Validity of the CALERIE 2 cohort—Comparisons were made to determine if there were systematic differences in demographic and psychological variables in volunteers who were excluded from clinic screening or dropped out from baseline visits compared to volunteers who were randomized to the intervention. Of note is that telephone screenings were not included in this analysis. Signed consents allowed collection of more refined data for analysis; thus, only volunteers who came to the first in-clinic screening visit and signed the corresponding informed consent were included in this comparison. Tests of group differences were assessed by Pearson Chi-square statistics for nominal variables, and by 2-sample Wilcoxon tests for continuous interval level variables.

Results

Recruitment

The recruitment sources for the initial phone screens across all of the sites are outlined in Table 2.

For PBRC, approximately 54 phone screens were needed to yield one randomization, thus, 4320 phone screens were needed to randomize 80 participants. The high ratio of phone screens per randomization is an indication of the level of difficulty to recruit for this trial. As a result of its broad catchment area, the most effective media of paid advertisement were

direct mail, newspaper (especially orange “sticky” notes on the front page), and television. Community events, radio, and the PBRC billboard were the least effective in yielding phone screens. Finally, the variability of quality between community events and radio stations made it difficult to utilize these media as effective methods of recruitment.

For Tufts, the most successful advertising media at Tufts were newspapers, flyers, subway advertisements and the CraigsList website. Least successful were health fairs/expos and mail outs (especially church mailings). At Tufts, 59 phone screens were needed for each randomized volunteer, and therefore 4248 calls were needed to complete the randomization of 72 participants.

For Washington University, recruiting via TV and print media worked well for Washington University. A total of 1714 phone screens were completed, and recruiting targets were met on a monthly basis. Thus, 25.2 phone screens were needed, on average, for each randomized volunteer, and the phone screens yielded 68 randomized participants overall.

Screening and Enrollment

CALERIE phase 2 screening began in April 2007, with the goal of randomizing 225 participants to a two-arm (CR or control) trial within 3.5 years. Recruitment ended February 2010. Two hundred, thirty-eight volunteers began baseline evaluations, of whom, 220 were randomized, and 218 started the assigned intervention. Figure 1 presents the CONSORT diagram (12) describing the number of people seen at each step of the CALERIE recruitment and screening process.

A total of 10,856 volunteers contacted the clinical sites. A large number of potential participants (N= 9,787, 90%) were screened out after the telephone screening and/ or for not meeting a basic eligibility criterion determined in-clinic, e.g. BMI requirement. One thousand, sixty-nine individuals participated in the in-person clinic screening process after the BMI eligibility criterion was met and 831 (78%) were screened out during this process. Further details regarding reasons for exclusion are provided below.

The three sites of the CALERE Phase 2 study each had the goal of randomizing approximately 75 participants to the two-arm trial. Thus, PBRC enrolled 80 participants, Tufts enrolled 72, and Washington University enrolled 68. A volunteer was considered “enrolled” when s/he signed the baseline informed consent document and began the baseline evaluations. 238 participants underwent baseline procedures, of which, 18 were ineligible/ dropouts. From this, a total of 220 completed baseline testing and were randomized, of which 2 did not start the intervention. Thus, a total of 218 (n= 153 women, n=67 men) participants started the assigned intervention (CR group or control group).

The projected recruitment number/rate (approximately 7.5 participants per month) for this study was determined by consensus of the principal investigators at each site. The 218 participants were randomized over a period of approximately 35 months with a mean randomization rate of 6.2 participants per month. The cumulative enrollment (signed baseline consent form and began baseline evaluations) and randomization plot in Figure 2 shows the number of participants starting the baseline procedures and participants ultimately randomized. The plot shows that the enrollment and randomization rates were relatively constant throughout the study.

Reasons for Exclusion/Dropout

Reasons individuals were screened out or excluded from the screening and baseline testing are summarized for the entire CALERIE study as well as by each clinical site in Table 3. Study-wide, the primary reasons for exclusion were: 1) BMI too high (21.7%), 2) too old

(13.5%), 3) personal commitments (13.0%), 4) changed mind (11.0%), 5) regular medication use (8.7%), and 6) BMI too low (8.6%). For the specific sites, the most frequent reasons for exclusion differed. At PBRC, individuals were most frequently excluded for BMI too high (32.2%), and age too old (20.7%). At Tufts, the top reasons included, changed mind (18.6%), personal commitments (18.3%), BMI too high (13.0%), and BMI too low (12.3%). For Washington University, the top reasons included, age too old (22.4%), personal commitments (21.2%), BMI too high (19.1%), and regular medication use (15.0%).

External validity of the CALERIE cohort

Participants who were randomized were compared to those who were excluded or screened out at any point during in- clinic screening or baseline visits. Comparisons were made with respect to age, sex, race, ethnicity, marital status, living situation, education level, family income, weight, BMI, BMA, MAEDS, BDI, and EI (Table 4).

Randomized participants were significantly older than those not randomized ($p < .0001$) and slightly heavier ($p = .0051$). A higher percentage of randomized participants were White ($p = .0006$), female ($p = .04$), married ($p < .0001$), had a college degree ($p < .0001$), and had higher income ($p < .0001$). In addition, the two groups differed very slightly on scores on the EI restrictive eating scale and the restrictive eating subscale of the MAEDS ($p < .05$).

Discussion

The CALERIE Phase 2 study is the first randomized controlled trial that systematically investigated sustained CR on aging in relatively healthy, non-obese humans. The strategies that were utilized for participant recruitment and the screening of candidates for the study were successful and enrollment goals were met. A large percentage of the initially interested volunteers were excluded during screening and this reflects the strategy of the CALERIE group to proactively eliminate volunteers before baseline who were ineligible or unlikely to adhere to the rigors of the study. Importantly, only 18 of the 238 volunteers (7.6%) dropped out during the entire baseline period (5= consent withdrawn, 3=adverse event, 10= other). Thus, if a volunteer passed the extensive screening process, it was likely that he/she would go on to complete the baseline procedures. The baseline period may also be considered a “boot camp” to further weed out individuals who would likely not complete the trial as well as record baseline data. The dropouts/exclusions in the actual baseline period were low. The higher exclusion numbers were included on the front end in screening, which you would expect in a study of this nature.

There were two main challenges in the recruitment and screening of the CALERIE Phase 2 study follow. First, recruitment of men posed a unique challenge. Achieving a 50:50 representation of men and women in the study was not feasible given the nature of the intervention and outcome evaluations (e.g. men presented with more work schedule conflicts). However, intensive advertising and recruitment campaigns that specifically targeted men were implemented at the clinical sites which yielded a 70:30 split of women and men that were randomized. Further, results from the Phase 1 studies suggested that it was unlikely that there would be gender-specific effects for either the primary and/or secondary outcomes. Thus, the unequal ratio of men and women in the study were not of high concern and could be addressed in the statistical analysis. Second, recruitment for racial/ethnic minorities posed a challenge. Significant effort was made to recruit an ethnically diverse group of individuals based on the demographics of the three clinical sites, and yielded a 24% representation of minorities as a result.

Overall, there were some differences in demographic and psychological variables between those randomized and those screened out of the CALERIE Phase 2 study. Individuals who

were randomized were primarily Caucasian, women, slightly heavier, and older. Individuals who were randomized had greater resources and/or support than their non-randomized counterparts, e.g. married, higher income, and more education. On psychological variables, randomized participants showed slightly higher restrictive eating scores on one measure of restrictive eating (EI), but slightly lower on another measure of restrictive eating (MAEDS subscale). Nonetheless, these differences were extremely small and not clinically meaningful. In sum, the randomized participants were slightly heavier (possibly more motivated for health improvement and weight loss) and had more flexibility with time/ schedule and resources to participate in the study.

Compared to other large intensive lifestyle change studies, the CALERIE Phase 2 study screened 10,856 volunteers and randomized 220 individuals (2.0%). In comparison, the Diabetes Prevention Program (DPP; 13), screened 158,177 individuals and randomized 3,819 (2.4%); Look AHEAD (Action for Health and Diabetes; 14) screened 28,622 and randomized 5,145 (18.0%); the Arthritis, Diet, and Activity Promotion Trial (ADAPT;15) screened 2,209 and randomized 316 (14%); and the Hypertension Prevention Trial (HPT; 16) screened 11, 810 and randomized 841 (7%). All of the aforementioned studies required specific inclusion/exclusion criteria and all employed fairly demanding study regimens and time investments on the part of the participants. The inclusion ratio in CALERIE was similar to that in the DPP study and lower than those in Look AHEAD, ADAPT and HPT. The variance in these studies is likely due the level of specificity/restrictive nature of the inclusion/exclusion criteria and/or the intensiveness of the requirements/time commitment for participation in the study.

It should also be noted that CALERIE Phase 2 was not a typical clinical trial, as the trials mentioned above. Typically, a clinical trial assesses the clinical impact of some intervention on a clinical or biologic outcome, with the goal of changing clinical care. This study did not aim to investigate weight loss or the efficacy of different weight loss modalities, nor was it about the impact weight loss in obese individuals. Rather, it investigated the physiological and psychological impact of sustained caloric restriction in normal weight and slightly overweight individuals, who would otherwise likely not require a weight loss or caloric restriction program. Thus, it was not a traditional clinical trial, but more consistent with a model of a controlled experiment in free-living humans. Other than assessment of these impacts to advance science, there was no medical or clinical advantage for the individual that was an objective of this study, which we believe contributes to the restrictive nature of the inclusion/exclusion rates of such a rigorous and time consuming study.

The recruitment and screening procedures in the CALERIE Phase 2 study were utilized in an effort to find the best possible candidates in each demographic location of the study that would be able to meet the requirements of the two-year study. As a result of the intense screening and recruitment process, the CALERIE study offered several lessons for researchers planning similar life style intervention trials. First, recruitment should be site-specific. It should be tailored to the population at that particular site, and it must adapt to changing circumstances and approaches (e.g., Craig's List, wellness talks). Second, a multi-disciplinary screening process should be applied to address medical, physical, psychological/ behavioral suitability of participants. Third, a multi-step screening process with simple criteria evaluated first, followed by more elaborate procedures, we believe, facilitates efficient use of study resources. Although it may take some time to work through a multi-step process, individuals who do not meet basic criteria are excluded early and do not participate in the more intensive and costly in-clinic procedures. Fifth, volunteers should be challenged, e.g., 14-day food diary, to assess if s/he can perform a significant task for the study. Finally, once screening is finished, potential candidates for the study should be

reviewed thoroughly with a multi-disciplinary review of the volunteer's potential to adhere successfully to the requirements of the study.

A limitation of the study includes a lack of a cost-effectiveness analysis. A cost effectiveness analysis would be a contribution to this paper. Each site adapted to resources, people and structures that were available at the site. However, unfortunately, the data were not collected to specifically address this question, and is therefore, beyond the scope of this paper.

In sum, for a trial of this nature (long term lifestyle change for normal weight and slightly overweight individuals), we achieved the highest yield possible via our screening methods. Thus, we were as efficient as you can possibly be within the confines of the study participant requirements (e.g. recruiting normal weight individuals with no medical or psychological issues AND could adhere to the rigors of the trial). Significant thought was put into the design and process with lessons learned from CALERIE Phase 1 for a nondisease based model. We utilized standard methods of recruiting, screening and enrolling; however, our process was more intense than similar disease-based clinical trials. At first glance, while the methods may appear inefficient (high throughput, effort, and cost of the intense process), the ultimate cost of having non-compliant subjects in a trial that is so invested in adherence to the intervention, is much greater. By the use of intense screening and baseline procedures, we were able to obtain highly dedicated subjects. Considering that individuals were healthy to begin with in this trial, there were extreme requirements of them to ultimately complete the trial. Thus, the cost of getting the "right" subjects was offset by the compliance and long term participation of those ultimately randomized and contributed to the ultimate success of the trial. We recommend that future trials of similar interventions should employ a cost effectiveness analysis of such intense procedures.

Acknowledgments

Funding

This research is supported by the National Institute on Aging at the National Institutes of Health (grant numbers U01AG022132, U01AG020478, U01AG020487 and U01AG020480).

We are deeply grateful to the individuals who volunteered for participation in the study.

References

1. Helibronn LK, et al. Effect of 6-month calorie restriction on biomarkers of longevity, metabolic adaptation, and oxidative stress in overweight individuals: a randomized controlled trial. JAMA. 2006; 295:1539–1548. [PubMed: 16595757]
2. Das SK, et al. Long-term effects of 2 energy-restricted diets differing glycemic load on dietary adherence, body composition, and metabolism in CALERIE: a 1-y randomized controlled trial. Am J Clin Nutr. 2007; 85:1023–1030. [PubMed: 17413101]
3. Racette SB, et al. One year of caloric restriction in humans: feasibility and effects on body composition and abdominal adipose tissue. J Gerontol A Biol Sci Med Sci. 2006; 61:943–950. [PubMed: 16960025]
4. Rochon J, et al. Design and conduct of the CALERIE study: Comprehensive assessment of the long-term effects of reducing intake of energy. Journal of Gerontology. 2011; 66A(1):97–108.
5. Taylor-Piliae RE, Norton LC, Haskell WL, et al. Validity of a new brief physical activity survey among older adults. Circulation. 2005:111.
6. Stunkard AJ, Messick S. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. J Psychosomatic Res. 1985; 29:71–83.

7. Anderson DA, Williamson DA, Duchmann EG, Gleaves DH, Barbin JM. Development and validation of a multifactorial treatment outcome measure for eating disorders. Assessment. 1999; 6:7–20. [PubMed: 9971879]

8. Kutlesic V, Williamson DA, Gleaves DH, Barbin JM, Murphy-Eberenz KP. The Interview for the Diagnosis of Eating Disorders IV: Application to DSM-IV diagnostic criteria. Psychological Assessment. 1998; 10:41–48.

9. First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured Clinical Interview for DSM-IV Personality Disorders, (SCID-II). Washington, D.C: American Psychiatric Press, Inc; 1997.

10. Beck AT, Beamesderfer A. Assessment of depression: the depression inventory. Mod Prob Pharmacopsych. 1974; 7:151–169.

11. Stewart TM, Allen HR, Han H, Williamson DA. Development of the Body Morph Assessment Version 2.0 (BMA 2.0): Tests of reliability and validity. Body Image. 2009; 6:67–74. [PubMed: 19244002]

12. Moher D, Schulz KF, Altman DG. for the CONSORT GROUP. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Intern Med. 2001; 134:657–62. [PubMed: 11304106]

13. The DPP Research Group. The Diabetes Prevention Program: recruitment methods and results. Controlled clinical trials. 2002; 23:157–171. [PubMed: 11943442]

14. The Look Ahead Research Group. Reduction in Weight and Cardiovascular Disease Risk Factors in Individuals with Type 2 Diabetes. Diabetes Care. 2007; 30(6):1374–1383. [PubMed: 17363746]

15. Miller GD, et al. The arthritis, diet, and activity promotion trial (ADAPT): design, rationale, and baseline results. 2003; 24:462–480.

16. Nemat O, et al. Recruitment in the hypertension prevention trial. 1989; 10:30S–39S.

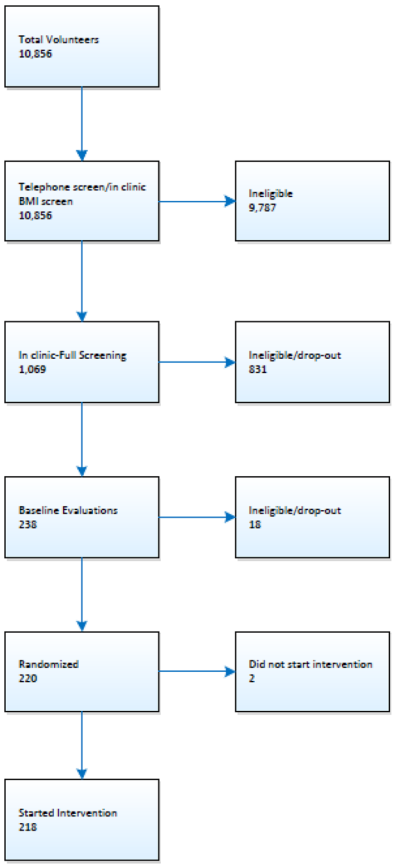

Figure 1.
The number of people seen at each step of the CALERIE 2 recruitment and screening process.

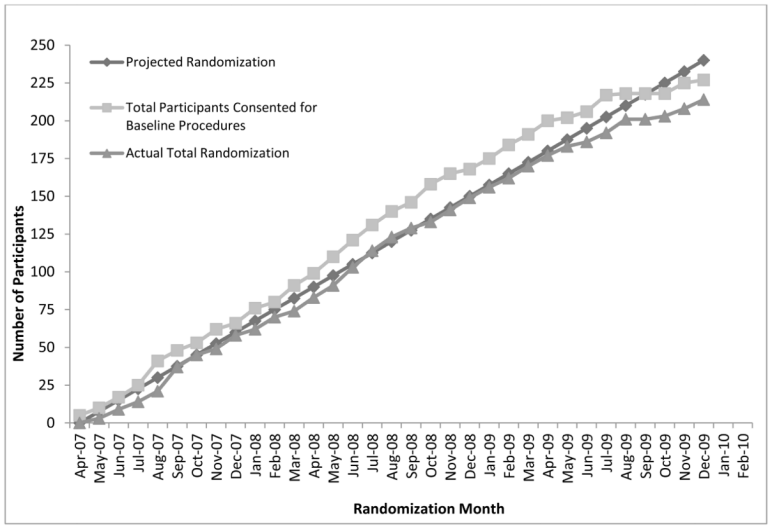

Figure 2.
Projected randomization, participants consented, and cumulative randomization of participants in the CALERIE 2 study.

Table 1

Screening procedures

Screening Visit	Activity	Description	Reason
Telephone Screening	Phone interview	• Trained recruiters familiar with the CALERIE protocol conducted an initial telephone screen. • After obtaining verbal consent, the recruiter completed a standardized form common to all sites that included: name, age (year of birth) height, weight, telephone number, address and basic eligibility information on demographics and health. • Candidates who met the basic eligibility criteria were scheduled for a clinic visit.	• Screen out clearly ineligible volunteers, e.g. wrong age, BMI, medical exclusions, before they came in for any in clinic screening visits
Screening Visit 1	• Consent • Verify BMI • Review information video • Demographic information • General health questionnaire • 7-day Stanford Physical Activity Record (PAR) • Eating Inventory (EI) • Multiaxial Assessment of Eating Disorder Symptoms (MAEDS) • Structured Clinical Interview for Diagnosis of DSM-IV Personality Disorders Questionnaire (SCID-II) • Beck Depression Inventory (BDI-II)	• A screening informed consent was signed by the participant and was documented on forms approved by the site's Institutional Review Board (IRB). • Body mass index (BMI, kg/m^2) was verified using measured height and weight. • Candidates watched a video presentation so they could understand the time and commitment that would be required for participation. The study coordinator was available to answer any questions. • Detailed demographic information pertinent to the study such as physician information, education level, race, ethnicity, occupation, marital status, emergency contacts etc. was obtained using a demographic questionnaire. • A Stanford brief physical activity questionnaire (5) was used to assess the activity level of each candidate to screen out those who are very physically active or are in training. • If indicated: candidates completed questionnaires and interviews to assess eating patterns, weight history, eating disorders, lifestyle and psychopathology. • Candidates met with the study psychologist/trained member of the behavioral team for a SCID-II interview (8) (i.e., given if the person met criteria for personality disorders based on the SCID-II Questionnaire), and the Interview for Diagnosis of Eating Disorders (IDED-IV; 10) (i.e., given if cutoff scores from the eating inventory or MAEDS were exceeded).	• Confirm simple eligibility criteria including BMI, demographics, physical activity, and some behavioral indices
Screening Visit 2	• Medical and Medications History	• Candidates were asked to come in to the center following a 12-hour overnight fast. A fasted clinic weight was measured at this visit.	• Confirm medical criteria, pregnancy, follow up on

Screening Visit	Activity	Description	Reason
	• Physical Examination • Standard 12- lead ECG • Blood chemistry, hematology, and urinalysis • Serum pregnancy test and contraception use (for women) • Barriers to Participation Interview • Body Morph Assessment 2.0 (BMA 2.0) • 14-day food record	• A study physician obtained a detailed medical and medications history and performed a physical examination to determine medical eligibility of the candidate. • A standard 12-lead electrocardiogram was recorded and reviewed by a physician and, if necessary, by a cardiologist. • Blood for hematology, clinical chemistry (including a lipid panel), a serum pregnancy test, where applicable, and urine for urinalysis was collected. A repeat blood draw was required if any abnormal laboratory values are found in the first sample, thereby initiating another clinic visit. • Barriers to participation in the study were assessed using a standardized interview which was administered by trained behavioral staff. This assessment ascertained participants' willingness to commit to the study, support from household members to participate in the study, motivation and challenges facing the participant and other similar study specific issues of importance. As part of this discussion, a template of the study schedule was provided and reviewed and the candidate was asked to complete a calendar of their work schedule and travel obligations. • A computer- based body image assessment test (11) was performed to identify candidates with body image concerns/perceptions that might indicate a risk factor for the development of eating disorder issues during the study. • A study dietitian used a study- specific diet questionnaire to obtain detailed information about the candidate's diet. Information was collected on major food allergies or special conditions (gastrointestinal conditions) which may affect participation in the study, current supplement use and feelings about discontinuing supplements for the study, alcohol consumption patterns and other diet related issues. • The dietitian also provided detailed verbal and written instructions on how to complete a 14-day food record. This process was utilized as a behavioral run-in and to assess the ability of a participant to adhere and complete a food record continuously over a 2 week period.	behavioral indices, and assess barriers to participation
Screening Visit 3	• Blood chemistry/ hematology/uri nalysis review - repeat blood tests if necessary and review at a follow-up screening visit,	• Candidates returned their completed 14-day food records and met with the dietitian to discuss the completeness, accuracy, and contents of the diet record. Issues such as underreporting, not reporting accurate portion size, not reporting recipes, and any missing information was discussed and, if	• Follow up and confirm good candidacy for study. Repeat medical tests as necessary, food record follow-up and follow up on

Screening Visit	Activity	Description	Reason
	• 14-day food record review, repeat if incomplete and review at a follow-up screening visit	necessary, a repeat food record was recommended. The candidate's willingness to reduce his/her dietary intake in the event that s/he was randomized to the CR group was also discussed. • Candidates met with study personnel to review work or personal schedule conflicts as well as regarding any test scores from questionnaires that required a more detailed discussion.	issues related to barriers to participation.

Table 2

Source of phone screens

	Number of Phone Screens		
Advertisement Medium	**PBRC**	**Tufts**	**Washington University**
TV	1674	98	265
Newspaper	837	1033	214
Mail out/brochure	578	58	—
Initial screening application/PBRC recruiter	330	—	—
Word of Mouth	320	—	—
Email/Website	311	110	317
Flyers	59	431	16
Billboard	45	—	—
Radio	45	65	63
BRGH Referral	8	—	—
Screening event / Presentation	7	27	—
Subway advertising	—	1377	—
Craig's list website	—	471	2
Local Magazine	—	140	5
Academic newspapers and intranet	—	177	—
Previous volunteer screening	—	—	95
Letters	—	—	399
"Volunteer for Health Program"	—	—	129
Referrals	—	257	119
Health fairs/expos	—	3	60
Internet searches	—	—	13
Site Employee Newspaper		—	22
No answer/unknown	175	551	178
Other	28	9	—
Total	**4417**	**4807**	**1897**

Note: — denotes "not applicable". Also, note that in Table 3, Washington University reports N=1632. The discrepancy of the N=1632 and the N=1897 in the above table is believed to be due to volunteers reporting multiple modes of media by which they came to inquire about the study.

Table 3

Reasons for exclusion

	PBRC		Tufts		Washington University		All Sites	
Total Number of Subjects Screened	**N 4417**	**pct^**	**N 4807**	**pct^**	**N 1632**	**pct^**	**N 10856**	**pct^**
Reasons for Exclusion								
Age - too young	53	1.2%	81	1.7%	2	0.1%	136	1.3%
Age - too old	898	20.7%	186	3.9%	350	22.4%	1434	13.5%
BMI too low	231	5.3%	580	12.3%	106	6.8%	917	8.6%
BMI too high	1398	32.2%	615	13.0%	298	19.1%	2311	21.7%
Medical exclusion criteria met	130	3.00%	184	3.9%	34	2.2%	348	3.3%
Abnormal lab values	0	0.00%	2	0.0%	2	0.1%	4	0.0%
Psychiatric & behavioral exclusion	50	1.2%	92	1.9%	10	0.6%	152	1.4%
Regular medication use	298	6.9%	397	8.4%	234	15.0%	929	8.7%
Prior Hepititus A/ pneumococcol vaccine	9	0.2%	37	0.8%	0	0.00%	46	0.4%
Personal commitments[1]	186	4.3%	865	18.3%	332	21.2%	1383	13.0%
Changed mind[2]	264	6.1%	882	18.6%	21	1.3%	1167	11.0%
Study concerns[3]	18	0.4%	122	2.6%	62	4.0%	202	1.9%
Protocol compliance concerns[4]	206	4.8%	202	4.3%	22	1.4%	430	4.0%
Met other exclusion criteria[5]	306	7.1%	399	8.4%	85	5.4%	790	7.4%
Unknown[6]	257	6.0%	87	1.8%	5	0.3%	349	3.3%
Other[7]	33	0.8%	4	0.1%	1	0.1%	38	0.4%
Total excluded	**4337**		**4735**		**1564**		**10636**	

^Percentage calculated based on the number of participants determined to be ineligible.

[1]Scheduling conflicts, will take too much time, lives too far from research site, transportation problems, child care issues

[2]Changed mind, lost interest in study

[3]Doesn't want to be involved in research, doesn't like/concerned about study procedures

[4]Unwilling/unable to comply with protocol requirements (randomization, CR intervention, discontinuation of dietary supplements and alcohol restrictions), unable to meet study demands

[5] Participated in CALERIE phase 1, excessive weight change, current or recent smoker, concurrent participation in another intervention study, current or planned pregnancy, breast feeding, heavy physical activity.

[6] Missed appointment, staff unable to contact subject, refused without explanation

[7] Reason for exclusion could not be classified in any other category.

Table 4

External validity of the CALERIE 2 cohort

Characteristic	Randomized N=220 Median (IQR)	Not Randomized N=849 Median (IQR)	p-value
Age	39.0 (32.2, 43.7)	34.9 (27.6, 42.1)	<.0001
Gender: (% Female)	153 (69.5%)	527 (62.1%)	0.04
Ethnicity (% Hispanic)	7 (3.2%)	52 (6.3%)	0.075
Race (% White)	168 (76.4%)	544 (64.1%)	0.0006
Marital Status (% married)	130 (59.1%)	322 (42.2%)	<0.0001
Education (% College Grad or more)	181 (82.3%)	504 (62%%)	<0.0001
Income (%>$40,000/yr)	188 (85.4%)	620 (73.7%)	<0.0001
Weight (kg)			
Overall	71.0 (66.0, 76.3)	70.1 (63.8, 77.4)	0.21
Male	79.5 (74.9, 86.7)	78.5 (72.3, 83.7)	0.037
Female	67.8 (63.4, 71.7)	65.5 (61.2, 70.8)	0.0044
BMI	25.1 (23.8, 26.4)	24.5 (23.2, 26.2)	0.0051
BDI-II	1.0 (0.0, 4.0)	1.0 (0.0, 5.0)	0.16
EI-Dietary Restraint	11.0 (8.0, 14.0)	10.0 (7.0, 13.0)	0.032
EI-Disinhibition	4.0 (2.0, 6.0)	4.0 (2.0, 6.0)	0.087
EI-Hunger	3.0 (1.0, 4.0)	3.0 (2.0, 5.0)	0.046
MAEDS-Depression	39.0 (35.0, 44.0)	40.0 (36.0, 46.0)	0.12
MAEDS-Binge Eat	46.0 (39.0, 53.0)	45.0 (39.0, 52.0)	0.35
MAEDS-Purge	45.0 (43.0, 48.0)	45.0 (43.0, 49.0)	0.05
MAEDS-Fear Fat	45.0 (39.0, 51.0)	45.0 (39.0, 53.0)	0.30
MAEDS-Restrict Eat	43.0 (40.0, 48.0)	44.0 (40.0, 50.0)	0.021
MAEDS-Avoid Foods	52.0 (47.0, 59.0)	52.0 (46.0, 60.0)	0.77
BMA-Current	53.8 (48.0, 59.5)	53.4 (46.6, 58.9)	0.24
BMA-Ideal	49.1 (43.8, 53.5)	49.0 (42.1, 54.6)	0.76
BMA-Dissatisfaction	4.7 (−1.9, 11.5)	4.3 (−2.3, 12.0)	0.68

Note: Due to missing values, the total number in the “not randomized” group varies from N=849, varying across outcomes. There was no missing data in the ‘randomized’ group. BMI=Body Mass Index, BDI-II= Beck Depression Inventory II, EI=Eating Inventory, MAEDS= Multiaxial Assessment of Eating Disorder Symptoms, BMA=Body Morph Assessment 2.0.