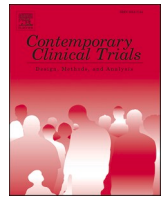




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



A multi-center trial of exercise and testosterone therapy in women after hip fracture: Design, methods and impact of the COVID-19 pandemic

Ellen F. Binder^{a,*}, Jesse C. Christensen^b, Jennifer Stevens-Lapsley^c, Jenna Bartley^d, Sarah D. Berry^e, Adrian S. Dobs^f, Richard H. Fortinsky^d, Kerry L. Hildreth^g, Douglas P. Kiel^e, George A. Kuchel^d, Robin L. Marcus^b, Christine M. McDonough^h, Denise Orwigⁱ, David R. Sinacore^j, Robert S. Schwartz^g, Elena Volpi^k, Jay Magazinerⁱ, Kenneth B. Schechtman^l

^a Division of Geriatrics and Nutritional Science, Washington University School of Medicine in St. Louis, St. Louis, MO, United States of America

^b Department of Physical Therapy and Athletic Training, University of Utah, Salt Lake City, UT, United States of America

^c Department of Physical Medicine and Rehabilitation, University of Colorado Anschutz Medical Campus, Denver, CO, United States of America

^d UConn Center on Aging, University of Connecticut, Farmington, CT, United States of America

^e Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, United States of America

^f Department of Medicine, Johns Hopkins Medicine, Baltimore, MD, United States of America

^g Division of Geriatrics, University of Colorado Anschutz Medical Campus, Denver, CO, United States of America

^h School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, United States of America

ⁱ Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, United States of America

^j Department of Physical Therapy, High Point University, High Point, NC, United States of America

^k Sealy Center on Aging, University of Texas Medical Branch at Galveston, Galveston, TX, United States of America

^l Division of Biostatistics, Washington University School of Medicine in St. Louis, St. Louis, MO, United States of America

ARTICLE INFO

Keywords:

Hip fracture
Sarcopenia
Frailty
Rehabilitation
Testosterone

ABSTRACT

Background: Up to 75% of hip fracture patients never recover to their pre-fracture functional status. Supervised exercise that includes strength training can improve functional recovery after hip fracture. The role of testosterone replacement for augmenting the effects of exercise in older women after hip fracture is unknown.

Methods: The Starting Testosterone and Exercise after Hip Injury (STEP-HI) Study is a 6-month Phase 3 multi-center randomized placebo-controlled trial designed to compare supervised exercise (EX) plus 1% testosterone topical gel, with EX plus placebo gel, and with enhanced usual care (EUC). Female hip fracture patients age ≥ 65 years are being recruited from clinical centers across the United States. Participants are community dwelling and enrolled within 24 weeks after surgical repair of the fracture. The EX intervention is a center-based program of progressive resistance training. The EUC group receives a home exercise program and health education. Participants receive dietary counseling, calcium and vitamin D. The primary outcome is the Six Minute Walk Distance. Secondary outcomes include physical performance measures, self-reported function and quality of life, and dual energy x-ray absorptiometry measures of body composition and bone mineral density.

Results: Enrollment, interventions, and follow-up are ongoing. We describe the impact of the coronavirus disease 2019 pandemic on the trial, including modifications made to allow continuation of the interventions and outcome data collection using remote video and audio technology.

Abbreviations: STEP-HI, Starting Testosterone and Exercise after Hip Injury; EX, supervised exercise; NIA, National Institute on Aging; COVID-19, coronavirus disease 2019; SMWD, six minute walk distance; EUC, enhanced usual care; EX+P, supervised exercise and topical placebo gel; EX+T, supervised exercise and topical 1% testosterone gel; CCC, Clinical Coordinating Center; DCC, Data Coordinating Center; sIRB, single Institutional Review Board; mPPT, modified Physical Performance Test; HIPAA, Health Insurance Portability and Accountability Act; SBT, Short Blessed Test; GDS, Geriatric Depression Scale Short Form; SPPB, Short Physical Performance Battery; ADL, activities of daily living; OARS, Older Americans Resources and Services; IADL, instrumental activities of daily living; BADL, basic activities of daily living; FSQ, Functional Status Questionnaire; MOP, Manual of Procedures; PT, physical therapy; DXA, dual energy x-ray absorptiometry; HRQ, Hip Rating Questionnaire; PROMIS®, Patient-Reported Outcomes Measurement Information System; CIPC, Central Imaging Processing Center; SHBG, serum hormone binding globulin.

* Corresponding author at: Division of Geriatrics and Nutritional Science, Washington University School of Medicine, 60 W. Euclid, Box 8303, St. Louis, MO 63110, United States of America.

E-mail address: ebinder@wustl.edu (E.F. Binder).

<https://doi.org/10.1016/j.cct.2021.106356>

Received 25 November 2020; Received in revised form 4 March 2021; Accepted 8 March 2021

Available online 11 March 2021

1551-7144/© 2021 Elsevier Inc. All rights reserved.

Conclusions: Results from the STEP-HI study are expected to have important clinical and public health implications for management of the growing population of hip fracture patients.

1. Introduction

Hip fractures represent a major public health problem of increasing magnitude. Annual incidence for women in the United States is approximately 789 fractures per 100,000 person-years [1], and an estimated 300,000 Americans aged 65 years and over sustain a hip fracture each year [2]. The majority of hip fractures occur in women over 60 years of age, and white women aged 85 years and older are the population at highest risk [3]. This is one of the fastest growing segments of our population, and therefore the burden of care for patients with hip fracture is expected to rise dramatically.

Although advances in clinical care have reduced mortality after hip fracture repair, many patients remain unable to achieve full functional recovery. Between 22% and 76% of patient with hip fracture do not recover to their pre-fracture ambulatory or functional status within 6 to 12 months of the fracture event [4–7], and 20% experience substantial loss of functional ability and independence [7,8]. Patients who experience such declines are more likely to be re-hospitalized or fall. Thus, despite receiving rehabilitation as part of their fracture management, many previously independent older patients require ongoing assistance with daily activities between 6 and 12 months post-fracture. Interventions designed to reduce physical impairments post-fracture may thus be beneficial for long-term recovery, functional independence, and as a pathway to reduce health care costs.

Several studies have demonstrated the benefits of structured exercise programs on functional outcomes in older patients with hip fracture although not all have shown changes that are clinically meaningful. Results from systematic reviews and meta-analyses also suggest that programs that include intensive strengthening after hospital discharge may improve short- and long-term mobility outcomes more robustly [9,10], but more evidence is necessary to support this conclusion [11], and to identify optimal exercise strategies leading to clinically meaningful improvements. The addition of an anabolic agent has potential to augment the effects of exercise on muscle function and thereby improve post-fracture function. To date, most studies of testosterone therapy conducted in women have focused on ameliorating osteoporosis or low libido [12–15]. Very few studies have evaluated the effects of anabolic steroids on skeletal muscle strength and physical function and most were conducted in younger, healthier women [16–19]. Although androgen levels decline with age in women [20–22], the impact of this change on age-associated declines in skeletal muscle mass and function is not well understood.

Recognizing that morbidity and poor functional recovery after hip fracture is a major public health problem in older women, the National Institute on Aging (NIA) awarded a 5-year grant to conduct the Starting Testosterone and Exercise after Hip Injury (STEP-HI) study. The STEP-HI study is a Phase 3 multicenter three-arm randomized controlled trial designed to compare a supervised center-based exercise program combined with 1% testosterone topical gel, with supervised exercise plus placebo topical gel, and with an enhanced usual care group, on changes in mobility in frail women after hip fracture. We describe here the methods of this study, including adjustments made to the protocol because of the coronavirus-2019 disease (COVID-19) pandemic.

2. Design overview and study hypotheses

STEP-HI is a Phase 3 multicenter three-arm randomized controlled trial that will enroll 168 women aged 65 years and older with a recent hip fracture. The primary objective of the STEP-HI trial is to test the hypothesis that supervised exercise combined with topical testosterone

therapy over a period of six months provides greater benefits for mobility as measured by the six minute walk distance (SMWD, primary outcome), than supervised exercise alone. A secondary objective is to demonstrate that six months of exercise combined with testosterone provides greater benefits for mobility over usual care after a hip fracture. Additional secondary objectives include assessment of changes in composite measures of physical performance, muscle strength, gait speed, patient-reported outcome measures of physical health and health-related quality of life, whole body and femoral neck bone mineral density, and total and appendicular lean and fat mass. A third exploratory aim is to obtain preliminary data regarding the sustainability of the effects of the exercise and topical gel interventions 3 months after discontinuation of the active interventions (i.e. 9 months after randomization) on self-reported measures of ADL function and quality of life.

Participants are randomly assigned to receive enhanced usual care (EUC), supervised exercise and topical placebo gel (EX+P), or supervised exercise and topical 1% testosterone gel (EX+T). A flow diagram of the study is provided below in Fig. 1.

Participants are being recruited at field centers in the United States located in St. Louis MO, Boston MA, Farmington CT; Baltimore MD, Denver CO, Galveston TX, Salt Lake City UT, and Pittsburgh PA. The Administrative, Clinical Coordinating Center (CCC), and Data Coordinating Center (DCC) and the single Institutional Review Board (sIRB) for the study are located at Washington University in St. Louis. The sIRB has approved the STEP-HI protocol and amendments. The NIA has appointed an independent data safety and monitoring board (DSMB) to oversee the trial.

3. Inclusion/exclusion criteria

Eligibility criteria for STEP-HI were designed to target older females (≥ 65 years) who: a) have undergone a recent surgical repair (within 22 weeks at screening) of a non-pathologic fracture of the proximal femur (including femoral neck or intracapsular, intertrochanteric, and subtrochanteric fractures); b) have a low serum total testosterone level defined as < 60 ng/dL and; c) are community-dwelling or in assisted living prior to the hip fracture event and anticipated to return to that location after hospital discharge or completion of rehabilitation; d) have persistent functional impairment defined objectively as a modified Physical Performance Test (mPPT) [23] score between 12 and 28. The mPPT range was chosen because in our previous exercise studies hip fracture patients with an mPPT score below 12 had difficulty completing research assessments and supervised exercise, while those with a score above 28 were functioning independently at a high level. Although many older patient with hip fractures reside in nursing homes, the overall goal of the study is to promote the continued independence of community-dwelling women after a hip fracture.

Study exclusion criteria (Table 1) were designed to identify those with safety risks related to testosterone therapy or to progressive resistance training, those who might have difficulty following directions or with adherence to study interventions, those who take medications that could interfere with the anabolic effects of testosterone, or who reside too far from the research center to be practical or financially feasible for the study. To maximize recruitment and the generalizability of the sample, the exclusion criteria were specified as liberally as possible, without jeopardizing safety. Each clinical site was responsible for establishing their geographical distance criterion, based on traffic patterns and transportation costs.

4. Recruitment and pre-screening

Participants are recruited from local hospitals, skilled nursing facilities, home care programs, orthopedic practices, and outpatient physical therapy practices, which are sites of care for patient with hip fracture during the immediate post-surgery recovery period. Because of Health Insurance Portability and Accountability (HIPAA) privacy regulations, research staff have access to electronic medical records only at affiliated institutions or facilities that have provided a HIPAA waiver to review specific data for purposes of recruitment. This has necessitated that STEP-HI staff perform extensive outreach to develop referral sources and maintain contact on a regular basis. To facilitate recruitment efforts, the study has a website (www.stephstudy.wustl.edu). Recruitment networks have varied by site, depending upon factors such as the strength of the relationships developed with orthopedists at hospitals not affiliated with the research site, and the level of interest by nursing home staff. Some clinical sites have contracted with hospitals to perform chart reviews and provide referrals of those patients who meet inclusion criteria. One site obtained permission for a HIPAA waiver to screen nursing home charts of patients admitted with a hip fracture. Another site created a Best Practice Advisory within the electronic health record of their healthcare system to alert study personnel of new patients with hip fracture. Publicity about the study has included articles in local newspapers and newsletters, televised interviews, posters, Facebook and Twitter posts from the accounts of sponsoring institutions. With permission from the patient, staff at referring facilities and clinical practices inform study staff of that person's interest in the study and provide contact information.

A prescreening questionnaire is administered either in person or by phone to those women who have agreed to be contacted about the study. The prescreening interview includes an explanation of the study, collection of demographic information, a limited medical and medication history, and evaluation of eligibility. If the person is eligible, an in-person visit is scheduled to explain the study in more detail. The visit is

conducted at the rehabilitation site or the participant's residence. Depending upon the participant's readiness to engage with the study, the consent form may be reviewed at that visit. A second visit is often necessary to review the participant's questions, discuss with interested family members, and review the consent form. All participants must provide their own informed consent; those who are unable to provide consent are excluded.

5. Study measures

A schedule of all study assessments is provided in [Table 2](#).

5.1. Screening

Women eligible based on the pre-screening interview are carefully screened to ensure that it is safe for them to participate in the assessments and interventions. Screening is not conducted until the participant is close to, or soon after, discharge from physical therapy and no later than 22 weeks after surgical repair of the fracture. Although it might be optimal to start the interventions early during the patient's recovery, we found that most women do not want to begin until after they have completed their prescribed therapies, due to physical demands and time constraints. Some of the screening assessments can be conducted at the rehabilitation site or the participant's home.

5.1.1. Cognitive screen

The Short Blessed test (SBT, range 0–28, higher score indicates greater cognitive impairment) [24] consists of seven cognitive tasks administered by study staff to evaluate for cognitive impairment, and is the first assessment administered. Women with a score ≥ 11 do not undergo further testing.

5.1.2. Demographics

Demographic information, including birthdate, sex, race, ethnicity,

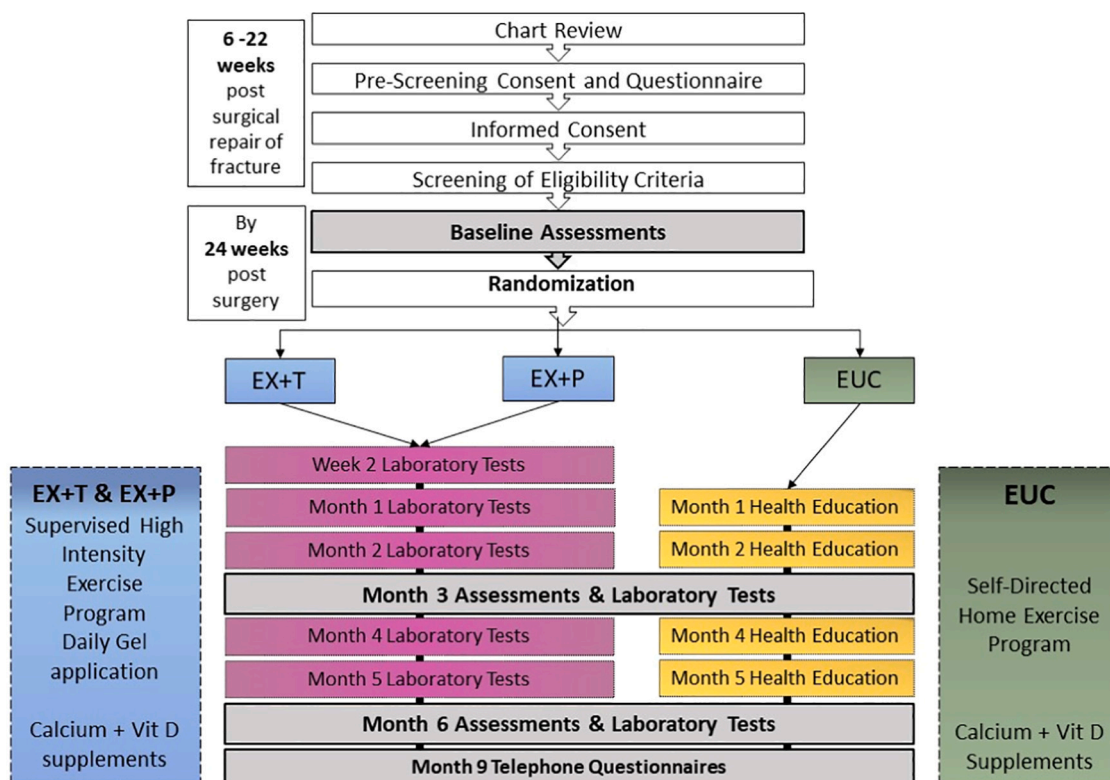


Fig. 1. Flow Diagram of STEP-HI Study Design

Table 1
Inclusion and exclusion criteria.

Inclusion criteria
<ul style="list-style-type: none"> Female 65 years and older. Surgical repair of a non-pathologic fracture of the proximal femur Surgical repair date that is within 22 weeks at the first screening visit Modified Physical Performance Score (mPPT) of 12–28. Serum total testosterone level < 60 ng/dL. Community-dwelling or in Assisted Living prior to the hip fracture.
Exclusion criteria
<ul style="list-style-type: none"> Cognitive impairment or dementia of severity sufficient to interfere with ability to provide informed consent or fully participate in the study, or a score of ≥ 11 on the Short Blessed Test of Orientation, Memory and Concentration. Use of progestin or androgen containing compound within the previous 6 months. Visual or hearing impairments that interfere with following directions for research procedures. Active or unstable cardiopulmonary disease (recent myocardial infarction, unstable angina, congestive heart failure class III or IV) within prior 6 months, which would limit full participation in this study. Uncontrolled hypertension, defined as a systolic BP > 160 mmHg or diastolic BP > 95 mmHg, on at least two occasions. Respiratory disease requiring chronic continuous oxygen therapy, or oxygen therapy during walking or exercise, which would limit full participation in this study. History of: a) Breast, ovarian, endometrial or cervical cancer with diagnosis within the previous 10 years; b) Breast, ovarian, endometrial or cervical cancer of Stage 2 or higher. Erythrocytosis with hematocrit $\geq 51\%$ at all sites except at Denver or Salt Lake City sites, or $\geq 52\%$ at the Denver or Salt Lake City sites. Severe anemia defined as Hgb < 7 g/dL. Uncontrolled diabetes defined as HgbA1C > 10%. Geriatric Depression Scale-Short Form score ≥ 12 at the screening assessment. Elevated liver transaminase or alkaline phosphatase levels ≥ 2.5 times above normal range. History of HIV or active viral hepatitis. Recent history of alcohol or substance abuse, or current alcohol intake of ≥ 10 drinks/week. Untreated or unstable thyroid disease, with serum TSH level ≥ 10 mIU/L or ≤ 0.4 mIU/L. Current use of aldactone, flutamide or leflunomide. Treatment with systemic corticosteroids (daily dose >5 mg prednisone or equivalent) for at least 90 days within the previous 12 months. Musculoskeletal or neurological conditions that limit full participation in this study, could be made worse by exercise training, or not expected to improve with exercise. End Stage Renal Disease on dialysis or GFR < 15 mL/min. History of idiopathic deep venous thrombosis or pulmonary embolus (i.e., not related to period or immobilization or surgery), recurrent or multiple venous thrombi; history of a hypercoagulable state such as Factor V Leiden thrombophilia. Lower extremity amputation other than toes. Planned joint surgery during the intervention period. Severe lower extremity pain or ulceration that could limit full participation in this study. Allergy to gel components. Residence too far from research center (specific distance to be determined by each site), or planned travel greater than 2 weeks within the next 9 months. Anticipated to be permanently living in a nursing home at the time of randomization. Site investigator's judgment that the participant would not be able to complete research procedures or interventions. Participation in another research study that in the site investigator's judgment could interfere or conflict with STEP-HI research assessments or interventions.

years of education, marital status, current living situation, and anticipated travel during the subsequent 9 months is obtained.

5.1.3. Medical evaluation

Standardized questionnaires are administered to obtain information about the participant's medical history, concomitant medications, and a review of systems. The Seattle Angina Questionnaire [25] is administered to assess for active symptoms of cardiac ischemia. The Geriatric Depression Scale (GDS, Short Form) [26,27] is administered to assess for symptoms of depression. Individuals with a score of ≥ 12 are excluded. A medically qualified professional performs a complete physical examination of the participant. A 12-lead electrocardiogram is performed and interpreted to assess for the presence of cardiac ischemia or arrhythmias.

5.1.4. Composite physical performance

The Modified Physical Performance Test (mPPT) is a 9-item standardized evaluation of physical function that is a modification of the PPT developed by Reuben and Siu [28]. It is associated with degree of disability, loss of independence, and mortality in older adults [29]. The mPPT substitutes the timed chair stand and standing balance tasks developed by Guralnik et al. (as part of the Short Physical Performance Battery, SPPB) for the writing and simulated eating items in the original PPT. In a previous study of exercise in frail hip fracture patients, the mPPT score significantly improved in response to the intervention [30]. For STEP-HI, all tasks are administered on site at the research center using standardized equipment at all sites (chair with 18 in. seat height), and the 4-m walk task is administered, so that composite scores can be calculated for both the mPPT and the SPPB.

5.1.5. Pre-fracture activities of daily living (ADLs)

To collect information about pre-fracture functional status, modified versions of the Older Americans Resources and Services (OARS) instrumental ADL (IADL) and basic ADL (BADL) questionnaires [31], and the Functional Status Questionnaire (FSQ) [32] are administered. An Assistive Device Checklist is also administered.

5.1.6. Non-fasting phlebotomy

If the participant meets eligibility criteria based on the SBT, mPPT, GDS and medical evaluation, blood samples are obtained for serum total testosterone level, hemoglobin A1C level, complete metabolic panel, and complete blood count. Women who are excluded because of a low mPPT score, high GDS score, or because of other medical exclusions that are anticipated to improve, can be screened again within the 22-week time period after surgical repair.

5.2. Baseline and follow-up assessments

Women who complete screening and meet eligibility requirements complete baseline assessments within 3 weeks prior to randomization. Assessments at 3 and 6 months from baseline are completed within a visit window of two weeks. The order of test administration, including sub-tasks for the mPPT and SPPB, is specified in detail in the Manual of Procedures (MOP) and is consistent at all visits.

5.2.1. Physical therapy assessment

Participants undergo a detailed musculoskeletal examination at baseline by a site Supervising Physical Therapist (PT). The purpose of this evaluation is to identify any physical limitations (joint deformities or range of motion limitations, hip precautions, gait instability, etc.) or musculoskeletal pain that might limit the participant's ability to perform any of the specified exercises. Information from this assessment is used by the Supervising PT to guide the site exercise interventionist in the implementation of the exercise program and, to avoid pain or injury.

5.2.2. Nutritional assessment

At baseline and Month 3, all participants undergo a one-hour nutritional assessment by a dietician or individual with training in nutrition, to: 1) document the participant's food intake, primarily focusing on the macronutrient composition of the diet, 2) calculate caloric and protein needs and, 3) evaluate for recent weight loss. Participants complete a food log for at least 3 days and up to 7 days during the period of the baseline and 3 month assessments. The dietitian reviews the log, discusses the information, and enters the most "typical day" into a web-based food record (HealthWatch 360™). Based on the food record, calculations of estimated protein (1.2 g/body weight in kilograms) and caloric needs (Mifflin St. Jeor Equation [33]), and based on the participant's expressed food preferences and budget, the dietician provides each participant with simple written recommendations at each time point to optimize their diet.

5.3. Primary outcome and secondary outcome measures

All outcome measures are collected at baseline prior to randomization, and at 3 months and 6 months, with the exception of dual energy x-ray absorptiometry (DXA) scans, which are only obtained at baseline and 6 months. Personnel conducting the testing for the primary and secondary outcome measures are blinded to treatment assignment.

5.3.1. Primary outcome - six minute walk distance (SMWD)

The SMWD is the primary study outcome and the first test performed at all blinded outcome assessment visits. The SMWD is a standardized, validated measure of mobility, physical function and cardiovascular fitness [34,35] with sensitivity and responsiveness to change in hip fracture patients [36]. The walking course is 25 m at all sites except at UConn Health where, for logistical reasons, it is 18 m. All sites use a standardized layout for the walking course with standardized cones and tape to delineate the course and turns. Participants are encouraged to use their usual assistive device for walking. The distance walked in 6 min is measured with a standardized stopwatch, lap counter, and measuring tape.

5.3.2. Secondary outcome measures

The mPPT and four-meter walk tests are administered as discussed above. Skeletal muscle strength of both lower extremities is measured using a standardized leg press machine (InFlight™ ILPC Model) at all clinical sites. Participants are oriented to the machine and the leg press exercise, and then perform a 1-repetition maximum test using both legs simultaneously.

Handgrip strength is measured with a hand-held dynamometer (Jamar, models vary by site) using a standardized protocol [37].

The Hip Rating Questionnaire (HRQ) [38] and the Patient-Reported Outcomes Measurement Information System (PROMIS®) Global Health questionnaire [39] are administered by study personnel to collect information about quality of life and pain. Self-reported information about ADLs is collected using standardized, validated questionnaires that measure difficulty with performance of 9 ADLs (Functional Status Questionnaire, FSQ), assistance with performance of 7 basic ADLs (OARS BADL scale), and 7 instrumental ADLs (OARS IADL scale) [40].

The GDS is administered to measure symptoms of depression. The Brief Resilience Scale is administered [41]. A standardized Assistive Device Checklist is administered to document all mobility devices and adaptive equipment being used by the participant at each time point.

Dual energy x-ray absorptiometry is performed using standardized procedures for whole body scans and contralateral femur scans. Images obtained from Hologic® scans are electronically transferred to a Central Imaging Processing Center (CIPC) for analysis. Images obtained from GE Lunar scans are reviewed by a designated CIPC staff member to confirm correct positioning and analyzed locally.

6. Safety assessments

All women undergo a screening mammogram at the start of baseline. Those randomized to the exercise and gel groups undergo a follow-up mammogram at 6 months, and those with a uterus also undergo a transvaginal or abdominal ultrasound at 6 months to assess for endometrial hyperplasia. A limited physical exam of all participants is performed by a medical provider at baseline, 3 months and 6 months, with a detailed assessment of body hair using the Ferriman Gallwey Scale [42]. Blood samples are obtained from all participants at baseline and 6 months for serum total and free testosterone, serum hormone binding globulin (SHBG), estradiol levels, 25-OH Vitamin D level, and serum lipid levels (non-fasting), and at 3 and 6 months post-baseline for complete metabolic panel and complete blood count. Women assigned to the exercise and gel groups provide blood samples at two weeks post-randomization and then monthly to monitor serum testosterone levels. For pragmatic reasons, non-fasting blood samples are obtained. For

women assigned to the exercise + gel groups, phlebotomy is performed 2 h after application of the gel by the participant. Whenever possible, staff observe the participant applying the gel in an effort to assess technique and adherence to gel application instructions. Circulating levels of testosterone in post-menopausal women are so low that there is no circadian rhythm and therefore it is not necessary to obtain an early morning specimen.

Transportation is provided to all study visits and exercise sessions, with taxis or a driving service. At one site, study staff provide transportation using a facility-owned vehicle.

7. Randomization

The STEP-HI Data Coordinating Center (DCC) has created an online password protected randomization system in REDCap to administer the random assignment of subjects to the three study arms. The study team member who has permission to randomize participants is assigned a unique, nontransferable user ID that is required to obtain random treatment allocations. To ensure that those performing evaluations remain blinded to treatment assignment, randomization is only performed by appropriate unblinded site investigational pharmacists who have received training and certification on STEP-HI randomization procedures. Following LOGIN, the group assignment is revealed only if all eligibility criteria are satisfied. To avoid temporal bias, randomization is blocked within clinical sites using random block sizes in order to preclude the possibility that investigators might know in advance the assignment of the last subject in a particular block. Participants are assigned to one of three treatment groups for 6 months:

1. **Enhanced Usual Care** (Home exercise + Health Education) (**EUC**)
2. **Exercise** (Supervised Exercise Training + Placebo gel) (**EX + P**)
3. **Combined** (Supervised Exercise training + Testosterone gel) (**EX + T**)

During the first 14 months of the study, eligible women were randomized to the study groups in a 1:1:1 ratio. Because of slow recruitment and a very low rate of participant withdrawals, in April 2020 the study DSMB approved a reduction in the target sample size and modification of the ratio of women assigned to the exercise and EUC groups, as discussed below in [Section 9](#).

8. Study interventions

8.1. Enhance usual care group (EUC)

Women assigned to the EUC group are prescribed a low intensity home-based exercise program that mimics standard clinical care at the time of discharge from physical therapy after a hip fracture repair. EUC participants are prescribed a standard set of 9 exercises (See [Appendix Table A1](#)) and instructed to perform three sets of 10 repetitions of each exercise 3 times per week (and are not discouraged from performing them more often). An in-person instructional session is provided by the site exercise interventionist, who demonstrates the exercises and instructs the participants in their performance. Written instructions with descriptive photos are provided at that time. The exercise interventionist meets with the participant monthly to review the exercises, observe their performance, provide guidance for correct performance, and provide standardized instructions for progressing some of the exercises. Participants maintain a standardized exercise log that is returned monthly. Adherence is measured as the percentage of exercise sessions completed out of a possible 72 sessions.

Participants in the EUC group also receive weekly contact by phone or email from the study staff to encourage adherence, answer questions, and maintain contact. They attend a monthly health education session to address health issues unrelated to exercise. The health education sessions are conducted by appropriately trained study staff, last 30–40 min,

Table 2Below provides a list of a complete **Schedule of All Study Assessments**.

Assessments	SC	BS	W2	M1 (W4)	M2 (W8)	M3 (W12)	M4 (W16)	M5 (W20)	M6 (W24)	M9 (W36)
Primary Outcome										
Six Minute Walk Distance		x				x			x	
Secondary outcomes (in the order of administration at baseline and follow-up visits)										
Leg Press 1-Repetition Maximum		x				x			x	
Hip Rating Questionnaire		x				x			x	x
Brief Resilience Scale		x								
Geriatric Depression Scale	x	x				x			x	
OARS ADL Questionnaires	x	x				x			x	x
Functional Status Questionnaire	x	x				x			x	x
Assistive Device Questionnaire	x	x				x			x	x
PROMIS Global Health		x				x			x	x
Handgrip strength		x				x			x	
Four Meter Walk		x							x	
mPPT	x	x				x			x	
Dual X-RayAbsorptiometry		x							x	
Additional measures										
Demographics	x									
Short Blessed Test	x									
Medical History										
Seattle Angina Questionnaire	x									
Medications	x	x	x	x	x	x	x	x	x	
Height		x				x			x	
Weight, pulse, blood pressure	x	x	x	x	x	x	x	x	x	
Physical exam	x	x				x			x	
Ferriman Gallwey Scale		x				x			x	
Electrocardiogram	x									
Mammogram		x							x*	
Dietary assessment		x				x				
Transvaginal Ultrasound									x*	
Blood samples	x	x	x*	x*	x*	x	x*	x*	x	

SC = Screening; BS = Baseline; W = Week; M = Month; mPPT = Modified Physical Performance Test.

* Only women assigned to the exercise groups.

and cover topics on healthy aging (excluding topics related to exercise or physical activity) with use of publicly accessible content using standardized materials prepared with presentation software. They coincide with the participant's monthly follow-up visit.

8.2. Supervised exercise program

Women in the EX+ P and EX+T study groups complete a supervised multimodal high-intensity exercise program that includes progressive resistance training. The exercise sessions are conducted on two non-consecutive days per week for 6 months at a dedicated exercise facility and directly supervised by an exercise interventionist, who is a physical therapist, exercise physiologist, certified personal trainer, or an individual with an appropriate level of experience, as approved by the study CCC. The supervised exercise program is a two-part progressively phased resistance-training program (See Appendix 1, [Tables A2](#) and [A3](#)). Phase 1 (baseline to end of month 1) consists of functional movements, stretching, balance, and strengthening exercises. Phase 2 (start of month 2 to end of month 6) consists primarily of higher-intensity strength training exercises. During specified exercises, the exercise interventionist targets the participant's eight-repetition maximum (8-RM), defined as the greatest resistance that can be moved 8 times through full range of motion in a controlled manner with good form and no movement substitutions. Supervised exercise participants are also expected to perform a home exercise program on three days of the week when they are not performing the exercise program at the clinical site facility. The home exercises include a progressive walking program (for a goal of 10–20 min total per day), and three of the exercises included in the supervised program, which vary over the course of six months.

8.2.1. Certification of exercise interventionists

The CCC Lead PT and staff train all site supervising PTs and exercise

interventionists in the supervised exercise and home exercise intervention procedures. Knowledge of procedures is tested by written examination and direct observations by the clinical site Supervising PT, who observes the site exercise interventionists conducting at least one Phase 1 session and at least one Phase 2 supervised exercise session with a volunteer (mock) participant using fidelity checklists for each phase. The sessions are also video recorded. The Phase 1 and Phase 2 fidelity checklist forms assess data completeness, physical performance, qualitative observations, and verbal and non-verbal communication. The checklist documents whether the participant completed the warmup exercises, necessary repetitions at the appropriate intensity and cool down exercises. If the checklist score is less than 90%, the Supervising PT proceeds with retraining and remediation until the exercise interventionist scores 90% or higher. When the exercise interventionist scores 90% or higher, the video file is uploaded via an encrypted computer server for the CCC Lead PT to review and score using the appropriate fidelity checklist. The CCC PT (or designated staff) review the Phase 1 and Phase 2 exercise session videos using fidelity checklists for each phase. If the score is 90% or higher, the exercise interventionist is considered "STEP-HI certified" and the CCC issues a certificate of completion. When a new staff member joins the project, training includes viewing of required training videos and the procedures described above.

8.2.2. Exercise intervention fidelity

For each newly certified exercise interventionist, the Site Supervising PT directly observes a total of four exercise sessions (Phase 1 and Phase 2, two of each) during the training of the first 2–3 study participants. The second observation of Phase 1 and Phase 2 is video recorded. If the exercise interventionist achieves $\geq 90\%$ on the fidelity checklist, the video recording is submitted to the CCC Lead PT for review. If the exercise interventionist scores $< 90\%$ a remediation plan is devised to

include refresher training to ensure accurate understanding of the protocol and follow-up observation visits. Once the remediation plan is completed, another fidelity checklist and video-recorded session is completed.

For ongoing fidelity of the protocol, the Supervising PT at each site directly observes the primary exercise interventionist (those who conduct ≥ 8 sessions/month) once per month and completes the Fidelity checklist. For alternate exercise interventionists (those who conduct < 8 sessions/month), the Supervising PT completes either a Phase 1 or Phase 2 fidelity checklist every 3 months or after 8–10 sessions have been performed, whichever occur first. Approximately every 6 months for primary exercise interventionists, and every 12 months for alternate exercise interventionists, the session is video-recorded. If the fidelity checklist score is $\geq 90\%$, the video is uploaded for review and scoring by the CCC Lead PT. Fig. 2 presents a flow diagram that summarizes the exercise intervention fidelity protocol:

8.2.3. Exercise adherence

Adherence to the supervised exercise program is measured as the percentage of exercise sessions successfully completed out of a possible 48 sessions. A successfully completed exercise session is defined as one in which 70% or more of the exercises are performed at the target number of repetitions. Exercise interventionists complete a checklist and exercise log at each visit documenting the number of sets, repetitions, weight used or other specifications of the exercise. If the participant can complete at least 5 repetitions, the set is considered complete.

8.3. Testosterone/placebo gel administration

Testosterone is provided in the form of a topical gel (generic 1.0% testosterone). Topical administration was chosen for STEP-HI because it provides continuous transdermal delivery of testosterone for 24 h. and it is easier to achieve and maintain stable physiologic levels of testosterone than intramuscular or oral preparations [43]. The potential advantages of topical gel over a transdermal patch include ease of delivery and dose

titration, better skin tolerability, and invisibility after application [44]. The placebo gel is a chemically identical except that it does not contain any testosterone. The active and the placebo gels are distributed to the clinical sites by the STEP-HI Central Pharmacy in identical pump-dispenser bottles (manufactured by Nemera, La Verpilliere, France), with labeling that allows the sites to maintain blinding of participants and clinical site staff. The site investigational pharmacy is responsible for receiving, storing, weighing, distributing, and destroying (after return) the testosterone and placebo bottles. To comply with United States Food and Drug Administration (FDA) requirements, the CCC monitors gel expirations semi-annually, and arranges for accuracy studies and a certificate of analysis.

Women assigned to the active treatment group (EX+T) are prescribed a dose of testosterone gel that is expected to achieve a slightly supra-physiologic serum testosterone level (target testosterone level 110–160 ng/dl; reference range 12–78 ng/dl). The initial gel dosage is 12.5 mg (1 pump depression per day of 1% testosterone gel). Serum testosterone levels are measured at 2 weeks and one month after baseline, and then monthly; dose adjustments are individualized by the CCC unblinded physician. Women with serum testosterone levels below the target range have their dose increased by 12.5 mg (1 pump depression) increments. Women with levels above the target range have their dose reduced to 12.5 mg (1 pump depression) every other day or less. To maintain blinding among all STEP-HI staff, each time an adjustment is needed for an EX+T individual, the CCC unblinded physician instructs an EX+P individual to adjust placebo dosage.

Participants meet with the study coordinator at the randomization visit to receive detailed oral and written instructions regarding the proper application of the gel. The study coordinator observes the participant self-apply the first dose of gel. Participants are educated about precautions to prevent other individuals from exposure to the gel, such as hand washing immediately after applying the gel and covering the application site with clothing after the gel has dried. Participants receive a standardized log form to record the skin location and the time of day of each gel application. Gel bottles are returned to the site

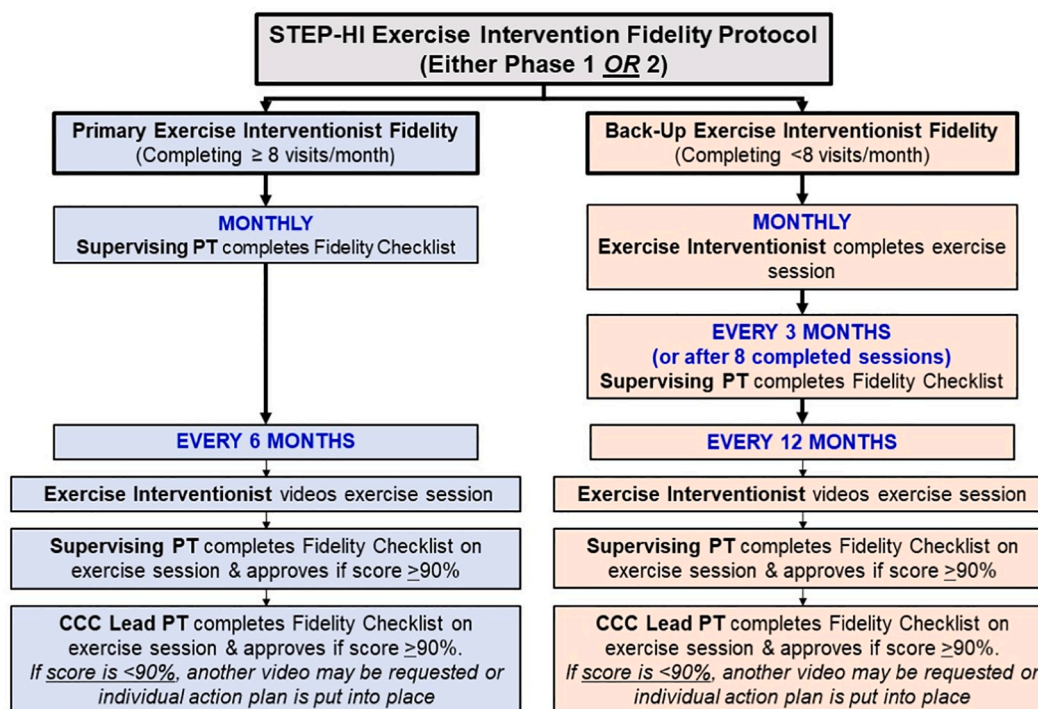


Fig. 2. Flow Diagram of Exercise Intervention Fidelity Protocol

pharmacy each month where they are weighed using a pharmacy or laboratory scale. Adherence to the gel is measured using bottle weights and serum testosterone levels.

8.4. Calcium and Vitamin D supplementation

Because of the high prevalence of Vitamin D deficiency in this population [45], and to minimize the potentially confounding effect of vitamin D myopathy, vitamin D supplements are provided to all women enrolled in the study in the form vitamin D3 capsules 2000 IU (Major Pharmaceuticals, Inc., Livonia, MI) to be taken daily. Participants are also provided with calcium carbonate 1000 mg daily in divided doses (Major Pharmaceuticals, Inc., or Geri-Care Pharmaceuticals, Corp., Brooklyn, NY). Calcium and/or vitamin D are not provided to those women with contraindications identified by the site physician. Calcium bottles are distributed monthly, Vitamin D3 bottles are distributed every 3 months; pill counts are performed on returned bottles.

9. Sample size considerations

The primary aim of the STEP-HI study is to compare the change from baseline to six months in SMWD in women assigned to the EX+P group with the corresponding change in the EX+T group. Prior data led us to base those computations in the original grant application on projected changes in SMWD of 42.7 ± 69.2 m in the EX+P group and 82.2 ± 92.5 m in the EX+T group.

We did not have preliminary data in women from which to directly estimate the magnitude of the projected differences in SMWD from combining topical testosterone therapy with the supervised exercise program. Data were available from a randomized, controlled pilot study conducted using the same eligibility criteria for frailty (mPPT score between 11 and 28) and nearly identical exercise procedures involving 25 frail hypogonadal (T level < 350 ng/dL) men ≥ 65 yrs. who were within 6–16 weeks post hip surgery. In that study, the EX+P group had an increase in SMWD of $+42.7 \pm 69.2$ m between baseline and 6 months, as compared to $+82.2 \pm 92.5$ m in the EX+T group. We assumed that the benefit of adding testosterone to exercise in women after hip fracture repair would be of similar magnitude as the observed benefit in the men's study. The sample size requirements were 68 per group for a power of 0.8, and 91 per group for a power of 0.9. After accounting for a 20% dropout rate, these sample sizes increase to 85 and 114, respectively. Our original plan was to randomize 100 subjects per group (total sample = 300), with statistical power of 0.86, for an alpha level of 0.05.

Recruitment for STEP-HI was initiated across the clinical sites between January and March 2019. Recruitment has been more difficult than anticipated, and by April, 2020, only 39 women had been randomized. However, the dropout rate was much less than anticipated such that STEP-HI did not have any study participants withdraw during that time period. Based on this experience, we performed revised power computations using assumed dropout rates of 5% and 10%. Study investigators were blinded to all outcome data when these revised computations were performed. Because the primary goal of the study involves comparisons between the EX+P and the EX+T groups, we also proposed a revised protocol that would change the randomization ratio such that going forward, more subjects would be randomized into the exercise groups than into the EUC group. Specifically, we proposed to the DSMB a revised randomization scheme with the goal to complete randomization with 70 subjects in the EX +P group, 70 in the EX+T group, and 28 in the EUC group (total $N = 168$). If the number randomized is 120, there will be 50 subjects in each exercise group and 20 in the EUC group.

Using the projected SMWD changes (42.7 ± 69.2 m in the EX+P group and 82.2 ± 92.5 m in the EX+T group) and assuming a) two-sided tests at the 0.05 level of significance and, b) the number of non-dropouts will be either 66 or 47 per group (5% dropout rate total N of 168 or 120) or 63/45 (10% dropout rate), we obtained the following results: For a

total N of 168, the power will be 0.79 for a 5% dropout rate and 0.77 for a 10% dropout rate. If $N = 120$, the power is 0.64 for a 5% dropout rate and 0.62 if the dropout rate is 10%. We emphasized to the DSMB that they are only marginally below the initially planned 0.8 when the sample size is 168. The reasons for this relatively high power despite a reduction in sample size from the originally proposed $n = 300$ is the imbalanced randomization that we proposed and the very positive results we have experienced with respect to dropout rates. In April 2020 this plan was approved by the study DSMB and NIA.

10. Statistical considerations

10.1. Analysis plan for primary study aim

The primary study aim addresses the question of whether EX+T will induce greater improvements in the SMWD relative to EX+P. Analytic strategies are based on the fact that all outcomes, including secondary outcomes, are continuous variables scheduled to be measured on at least three occasions. Initial analyses will include chi square tests and *t*-tests to compare baseline values of key demographic, clinical, and physiologic parameters.

The primary analysis will be a mixed model repeated measures analysis of variance that evaluates statistical contrasts to determine if the change from baseline to 6 months for EX+T are significantly different compared to EX+P. These models will also evaluate whether there are significant between-group differences in the change from baseline to 3 months. Additional analyses of covariance will be performed that adjust for baseline values and covariates that differed across groups at baseline.

We selected mixed model repeated measures analysis of variance for between group comparisons because we expect the primary outcome (SMWD) to be normally distributed or transformable to normality. Model development will be preceded by an evaluation of the data, to determine the appropriate covariance structure. Past experience suggests that correlation coefficients between time points will be greater for time points that are closer together than for time points that are further apart. Therefore, we anticipate that an autoregressive or a Toeplitz covariance structure will be appropriate for these analyses. If correlation coefficients are approximately constant regardless of the time between assessment points, we expect that a compound symmetric covariance structure will fit the data best. Schwarz-Bayesian criterion and Akaike's Information criterion will be used to inform covariance structure selection. All other things being equal, we will give preference to correlation structures that require the estimation of the smallest number of parameters.

10.2. Analysis of secondary and tertiary aims

The mixed model approach discussed above will also be applied to the secondary outcomes. The appropriateness of the selected analytic methods will be assessed by determining whether the required distributional assumptions are satisfied.

Specific Aim 3 addresses the question of whether the effects of the interventions persist after three months of discontinuing them. Analyses in this aim will be similar to those described for the primary specific aim. The central differences are that analyses for this aim will include the 9 month time point and will focus on the statistical contrast that performs a between-group comparison of changes from 6 to 9 months for EX+T relative to EX+P. In addition to *p*-values, we will generate confidence bounds that quantify the changes from 6 to 9 months within each group and the potential differences between the 6 to 9 month changes in different groups.

Missing data will be a particular concern in our mixed model analyses. To minimize the burden of missing data, all subjects who withdraw from the study are encouraged to undergo the final assessments that, at the very least, provide data for the SMWD. Our analytic plan

focuses on determining whether data are missing at random. If so, we will consider missing data to be ignorable and will employ multiple imputation using PROC MI in SAS conduct standard mixed model analyses incorporating these imputed values. If we believe that data may be non-ignorably missing, we will adopt a conservative approach to analysis and impute values that bias towards the null hypothesis. For example, if evidence suggests that subjects may have dropped out for reasons related to limited response to the interventions, we will bias towards the null by assuming that the missing data on these subjects are equal to the worst value of the outcome measure that was observed for any subject at the time points that are missing. Finally, when we suspect non-ignorable missingness, we will perform sensitivity analyses that use more than one approach to imputation in order to evaluate the robustness of our conclusions.

11. Study procedures implemented during the COVID-19 pandemic

The COVID-19 pandemic placed considerable challenges on study recruitment and necessitated the rapid adaptation of procedures while maintaining study rigor and fidelity. Between late March 2020 and July 2020, recruitment at all STEP-HI clinic sites was halted due to research restrictions implemented in response to the COVID-19 pandemic. Personnel policies varied among the clinical sites, such that some allowed research staff limited access to research facilities (primarily to obtain supplies and process phlebotomy specimens), while a few restricted all access for staff and investigators. Women enrolled in the trial during this time were allowed to continue their assigned interventions if they could be safely conducted and monitored. Women in the EUC group were encouraged to continue the home exercise program and were contacted by the study coordinator monthly. Women in the supervised exercise groups were instructed to continue using their prescribed gel unless the STEP-HI Safety Officer determined, based on review of available testosterone levels and considerations such as inconsistent levels in the target range, lack of access to phlebotomy or inability to process specimens, that continued gel use would pose a safety risk. The supervised exercise program was innovatively adapted to be conducted using remote video technology, telephone or, in one case, by an appropriately trained family member when the participant was unable to manage the remote technology. A set of elastic resistance bands, dumbbells, and ankle weights were provided to exercise group participants. As a safety precaution, some of the balance exercises were eliminated. Exercise interventionists documented performance of the exercises and number of repetitions using study forms in use before the pandemic. At one site where the exercise interventionist was not available due to the pandemic, the exercise session was supervised remotely by a study interventionist located at another site. Participants returned on site as soon as the individual research site allowed. At sites where staff were not allowed on site, outcome data at the 3, 6, and 9-month time points were collected remotely. At a minimum, the blinded outcome assessor administered the Hip Rating Questionnaire, ADLs and QOL questionnaires by telephone.

Additional challenges encountered as a result of the pandemic include implementing procedures for personal protective equipment (PPE) during study procedures, increased time to screen patients for COVID-19, having participants wear a mask during exercises, and related costs which were not budgeted for originally. Sites are following institutional policies and procedures for screening and PPE, which vary slightly.

In the event that research restrictions must be implemented again in the future due to the pandemic, we will provide exercise equipment and tablets to participants for remote exercise sessions. Exercise sessions will

be recorded and reviewed by the supervising PTs and CCC PT for fidelity monitoring. A detailed remote exercise protocol has been written and exercise interventionists at all sites have been trained in anticipation of possible future COVID-19 shutdowns.

12. Discussion

The STEP-HI study is a Phase 3 randomized controlled trial that will evaluate the effects of adding testosterone therapy to intensive exercise on measures of physical function and recovery after hip fracture in frail female patients. The STEP-HI study features many innovations that are worthy of note. First, the recruitment strategy is uniquely tailored in each of the clinical sites to seek out the most productive and efficient ways of identifying potential participants. For example, many sites have aligned with home care agencies who are caring for the hip fracture patients. Others have created a network across the orthopedic practices in a geographic area surrounding the study sites and receive direct referrals to the study. One center is located in two post-acute care sites where women receive rehabilitation following a hip fracture. Another center created Best Practice Advisories within their healthcare system EHR to alert study staff of new hip fractures. Direct hospital-based recruitment is limited by the brief lengths of stay typical of patient with hip fracture and the precarious physical and cognitive status immediately following surgery. In some sites, alignment with surrounding skilled nursing facilities is advantageous.

A second innovation of the STEP-HI study involves the intervention design itself. The exercise intervention has been carefully designed to take into account the gradual return to full mobility inherent in recovery from fracture and surgery. Thus, the two phases of exercise over the six-month intervention period is carefully led by trained exercise interventionists who were trained uniformly and have had their performance evaluated for fidelity using video monitoring on remote devices. The testosterone gel management is individualized by measuring testosterone concentrations and dose adjustments to achieve a targeted level in the active arm accompanied by blinded alterations in the placebo arm.

In addition to the testosterone replacement intervention, the STEP-HI study has other innovative features. Demonstrating treatment fidelity is an important methodological requirement for any study evaluating the efficacy of an intervention. Multi-modal exercise programs such as the STEP-HI protocol are especially challenging for fidelity measurements because of the importance of consistency movements for flexibility and resistance exercises, compounded by the challenge that frail patients with hip fractures may require modifications of some exercises due to physical impairments. STEP-HI has the additional challenge of ensuring treatment fidelity across clinical sites. To our knowledge, STEP-HI is one of the first exercise intervention trials utilizing video technology and detailed checklists to ensure consistency of exercise sessions across sites and interventionists, and over the duration of the trial. The STEP-HI team has developed methods that can be applied to other multi-site trials of exercise, particularly those conducted in frail patients and/or rehabilitation settings, including those with pragmatic or adaptive designs. This aspect of the trial has also allowed us to adapt to the challenges of managing a complicated multi-site intervention trial during the COVID-19 pandemic.

The design and conduct of the STEP-HI study exemplifies some of the challenges inherent in trials involving frail older adults. The experienced team anticipated many of these challenges such as recruitment, retention, safety, participant burden, intervention fidelity, and budget cuts by the funding agency. Recruitment started slowly after single IRB approval, study personnel training, and DSMB approval. Study sites took some time to fully develop their referral and screening network within

the geographic location of the centers. This delayed early recruitment. Many women who were approached during the recruitment were contending with their own well-being following their surgery and recuperation, and did not feel they could further extend themselves to complete exercise sessions over six months. Reluctance to taking testosterone deterred other women. On the other hand, many women were dissatisfied with the trajectory of their recovery and welcomed the potential benefits of personalized exercise training after the completion of their post-surgery rehabilitation. In many instances, the children of patients were strong supporters of their mother's applying this level of attention to recovery. As the target population included many women with disqualifying exclusion characteristics, study staff had to screen many frail older women to identify those who were eligible and willing to participate.

After the first six months of recruitment, the study leadership recognized that the target sample size could not be achieved with the enrollment rate. At that stage, with input from the lead statistician, a study re-design was undertaken and the power was recalculated based on the reduced attrition observed in the early stages and after the decision to allocate less women to the control arm, as the primary hypothesis was based on the EX+T versus EX+P comparison. This decision took advantage of an unanticipated success with participant retention that should be considered in other similar trials.

Further threats to the success of the study were imposed by the COVID-19 pandemic that shut down clinical research for months, posed life-threatening consequences to older adults, and challenged the management of study sites in terms of uniquely trained personnel whose job security was in jeopardy. The study quickly developed an approach to remotely supervising enrolled participants at home to perform exercises in substitution for the in-center program until research was resumed during the summer months of decreasing coronavirus cases. Remote phlebotomy services were used when possible, further sustaining fidelity to the study protocol. Some home visits were conducted to measure study outcomes at the participant's home.

When completed, the STEP-HI trial will be the largest and longest study of testosterone therapy in older women focused on optimizing muscle strength and related functional outcomes. Therefore, it has important implications for clinical practice and public health related to hip fracture management. It will answer an important knowledge gap related to the clinical significance of testosterone deficiency in frail older

women. Information obtained from STEP-HI will also provide important guidance for future research and clinical interventions in older women with physical frailty and sarcopenia related to mobility disability.

Funding

This work was supported by the National Institutes of Health grant number R01 AG051647.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We want to recognize the outstanding dedication and work of the participants and study staff at all the clinical sites and especially the STEP-HI Lead Study Coordinator, Kelly Monroe, MSW. The COVID-19 pandemic has posed enormous challenges for conduct of the trial, which they have all managed with great determination and grit.

STEP-HI study staff by site

Washington University in St. Louis: Peter Dore, MS; Mary Banach, BS, RD; Courtney Willoughby, MS; Megan Baldenweck, BS; Davvan Butler BS, ACSM-CPT; Elizabeth Beach, BS, RD.

Hebrew SeniorLife: Evelyn O'Neill, BS; Mary Jeznach-Vierling, RN; Danette Carroll.

UConn Health: Heather McAbee-Sevick, MS; Deborah Noujaim, MPH; Esmeralda Korkutovic, BS.

University of Colorado-Denver: Rachel Watkins, B.S.; Brandy Cuelar, B.S.; Jere' Hamilton, B.A.

University of Texas Medical Branch: Roxana Hirst, MS, CCRP; Rae Kretzmer, MS, RD, ACSM-CPT; Eloisa Martinez, BS, CCRP; Adetutu Odejimi, NP.

University of Maryland-Baltimore: Mary Louise Fine, BSN, RN; Christine Wade, BA;

Elizabeth Parker, PhD, RD; Chad Wessinger, MS, RD, LD, CSCS.

Appendix. Appendix 1

Table A1
Enhanced usual care (EUC) home exercises.

Home Exercise Program		# Repetitions (or time) per set	# Sets
Sitting	Sitting Chin Tucks (Neck Retraction)	10	3
	Sitting Trunk Twists (Left and Right)	10	3
	Trunk Side-Bending (Left and Right)	10	3
	Lateral Arm Raises	10	3
Standing	Scapular retraction	30 s	1
Floor exercises	Cat/Camel Stretch	10	3
	Prone Contralateral Arm/Leg Lifts	10	3
	Side-lying Hip Abduction	10	3
	Supine Bent Knees Fall Out Stretch	30 s (per side)	1

Table A2
Supervised exercise program: Phase 1 exercises.

Phase I		Number of Repetitions (or time) per set	Number of Sets
Warm up	Walking on Treadmill or Track	5 min	1
	Arm Circles	30 s/direction	1
Sitting	Lateral Arm Raises	8 RM	2
	Bicep Curls	8 RM (per side if separated)	2
Standing	Tricep Kickbacks or Triceps Pushdowns	8RM (per side)	2
	Wall Slides	8 reps	2
	Toe Walking	30 s	1
	Heel Walking	30 s	1
	Tandem Walking	30 s	1
	Standing Feet Together OR On Each Leg	30 s (per side if separated)	1
	Weight Shift Onto Single Limb	30 s (per side)	1
	Plantar Flexion (unilateral or bilateral)	8RM (per side if separated)	2
Floor exercises	Cat/Camel Stretch	30 s	1
	Side Lying Hip Abduction (each side)	8RM (per side)	2
	Quadricep Stretch	30 s	1
	Supine Hamstring Stretch	30 s	1
	Supine Bent Knees Fall Out Stretch	30 s (per side)	1
	Supine Lower Abdominal Exercise	30 s	1

Table A3
Supervised exercise program: Phase 2 exercises.

Phase II		Number of Repetitions (or time) per set	Number of Sets
Warm up	Walking on Treadmill or Track	5 min	1
Resistance Exercises	Choose 2 of the 4 upper body exercises per week and alternate:		
	Lateral Arm Raises	8RM	2
	Bicep Curls	8RM (per side if separated)	2
	Tricep Kickbacks or Tricep Pushdowns	8RM (per side)	2
	Seated Row (or Bent Over Row)	8RM (per side if separated)	2
Perform all the lower body exercises:			
	Plantar Flexion (unilateral or bilateral)	8RM (per side if separated)	3*
	Hip Extension (each side)	8RM (per side)	3*
	Side Lying Hip Abduction (each side)	8RM (per side)	3*
	Sit-to-Stands with Weighted Vest	8RM	3*
	Mini Squats with Weighted Vest	8RM	3*
	Tandem Walking	30 s	1
Machine Exercises			
	Leg Press	8RM	3*
	Leg Extension	8RM	3*
	Knee Flexion	8RM	3*

* Only two sets performed during Week 1 of Phase 2.

References

- [1] J. Swayambunathan, et al., Incidence of hip fracture over 4 decades in the Framingham heart study, *JAMA Intern. Med.* 180 (9) (2020) 1225–1231.
- [2] G.E. Lewinnek, et al., The significance and a comparative analysis of the epidemiology of hip fractures, *Clin. Orthop.* 152 (1980) 35–43.
- [3] S.R. Cummings, et al., Risk factors for hip fracture in white women, *N. Engl. J. Med.* 332 (1995) 767–773.
- [4] S.R. Cummings, S.L. Phillips, M.E. Wheat, Recovery of function after hip fracture: the role of social supports, *J. Am. Geriatr. Soc.* 36 (1988) 801–806.
- [5] A.M. Jette, et al., Functional recovery after hip fracture, *Arch. Phys. Med. Rehabil.* 68 (1987) 735–740.
- [6] K.J. Koval, et al., Dependency after hip fracture in geriatrics patients: a study of predictive factors, *J. Orthop. Trauma* 10 (1996) 531–535.
- [7] J. Magaziner, et al., Predictors of functional recovery one year following hospital discharge for hip fracture: a prospective study, *J. Gerontol. Med. Sci.* 45 (1990) M101–M107.
- [8] J. Magaziner, et al., Recovery from hip fracture in eight areas of function, *J. Gerontol. Med. Sci.* 55 (2000) 498–507.
- [9] H.H. Handoll, C. Sherrington, J.C. Mak, Interventions for improving mobility after hip fracture surgery in adults, *Cochrane Database Syst. Rev.* 3 (2011), CD001704.
- [10] J. Diong, N. Allen, C. Sherrington, Structured exercise improves mobility after hip fracture: a meta-analysis with meta-regression, *Br. J. Sports Med.* 50 (6) (2016) 346–355.
- [11] S.Y. Lee, et al., Effect of lower-limb progressive resistance exercise after hip fracture surgery: a systematic review and meta-analysis of randomized controlled studies, *J. Am. Med. Dir. Assoc.* 18 (12) (2017) 1096.e19–1096.e26.
- [12] E. Barrett-Connor, et al., A two-year, double-blind comparison of estrogen-androgen and conjugated estrogens in surgically menopausal women. Effects on bone mineral density, symptoms and lipid profiles, *J. Reprod. Med.* 44 (12) (1999) 1012–1020.
- [13] A. Kenny, et al., Effects of transdermal testosterone on bone and muscle in older men with low bioavailable testosterone levels, *J. Gerontol. Med. Sci.* 56A (2001) M266–M272.
- [14] N.B. Watts, et al., Comparison of oral estrogens and estrogens plus androgen on bone mineral density, menopausal symptoms, and lipid-lipoprotein profiles in surgical menopause, *Obstet. Gynecol.* 85 (4) (1995) 529–537.
- [15] S.R. Davis, et al., Testosterone for low libido in postmenopausal women not taking estrogen, *N. Engl. J. Med.* 359 (19) (2008) 2005–2017.

- [16] J.C. Lovejoy, et al., Exogenous androgens influence body composition and regional body fat distribution in obese postmenopausal women: a clinical research center study, *J. Clin. Endocrinol. Metab.* 81 (1996) 2198–2203.
- [17] S. Davis, K. Walker, Effects of estradiol with and without testosterone on body composition and relationships with lipids in postmenopausal women, *Menopause* 7 (2000) 395–401.
- [18] A. Dobs, et al., Differential effects of oral estrogen versus oral estrogen-androgen replacement therapy on body composition in postmenopausal women, *J. Clin. Endocrinol. Metab.* 87 (2002) 1509–1516.
- [19] G. Huang, et al., Testosterone dose-response relationships in hysterectomized women with or without oophorectomy: effects on sexual function, body composition, muscle performance and physical function in a randomized trial, *Menopause* 21 (6) (2014) 1–11.
- [20] R.A. Lobo, Androgens in postmenopausal women: production, possible role, and replacement options, *Obstet. Gynecol. Surv.* 56 (6) (2001) 361–376.
- [21] H.G. Burger, et al., A prospective longitudinal study of serum testosterone, dehydroepiandrosterone sulfate, and sex hormone-binding globulin levels through the menopause transition, *J. Clin. Endocrinol. Metab.* 85 (2000) 2832–2838.
- [22] S.L. Davison, et al., Androgen levels in adult females: changes with age, menopause, and oophorectomy, *J. Clin. Endocrinol. Metab.* 90 (7) (2005) 3847–3853.
- [23] M. Brown, et al., Physical and performance measures for the identification of mild to moderate frailty, *J. Gerontol. Med. Sci.* 55A (6) (2000) M350–M355.
- [24] G. Blessed, B. Tomlinson, M. Roth, The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects, *Br. J. Psychiatry* (1968) 797–811.
- [25] J.A. Spertus, et al., Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease, *J. Am. Coll. Cardiol.* 25 (2) (1995) 333–341.
- [26] G. Sheikh, J. Yesavage, Geriatric depression scale (GDS): recent evidence and development of a shorter version, in: T.L. Brink (Ed.), *Clinical Gerontology*, Haworth Press, New York, 1986, pp. 165–173.
- [27] J. Yesavage, T.L. Brink, Development and validation of a geriatric depression screening scale: a preliminary report, *J. Psychiatry Res.* 17 (1983) 37–49.
- [28] D.B. Reuben, A.L. Siu, An objective measure of physical function of elderly outpatients, *J. Am. Geriatr. Soc.* 38 (1990) 1105–1112.
- [29] D.B. Reuben, A.L. Siu, S. Kimpau, The predictive validity of self-report and performance-based measures of function and health, *J. Gerontol. Med. Sci.* 47 (1992) M106–M110.
- [30] E. Binder, et al., Effects of extended outpatient rehabilitation after hip fracture, a randomized controlled trial, *JAMA* 292 (2004) 837–846.
- [31] G.G. Fillenbaum, M.A. Smyer, The development, validity, and reliability of the OARS multidimensional functional assessment questionnaire, *J. Gerontol.* 36 (1981) 428–434.
- [32] A.M. Jette, P.D. Cleary, Functional disability assessment, *Phys. Ther.* 67 (1987) 1854–1859.
- [33] M.D. Mifflin, et al., A new predictive equation for resting energy expenditure in healthy individuals, *Am. J. Clin. Nutr.* 51 (2) (1990) 241–247.
- [34] P. Enright, et al., The 6-min walk test: a quick measure of functional status in elderly adults, *Chest* 123 (2) (2003) 387–398.
- [35] S. Lord, H. Menz, Physiologic, psychologic, and health predictors of 6-minute walk performance in older people, *Arch. Phys. Med. Rehabil.* 83 (7) (2002) 907–911.
- [36] M.A. Auais, O. Eilayyan, N.E. Mayo, Extended exercise rehabilitation after hip fracture improves patients' physical function: a systematic review and meta-analysis, *Phys. Ther.* 92 (11) (2012) 1437–1451.
- [37] H.C. Roberts, et al., A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach, *Age Ageing* 40 (4) (2011) 423–429.
- [38] N.A. Johanson, et al., A self-administered hip-rating questionnaire for the assessment of outcome after total hip replacement, *J. Bone Joint Surg.* 74A (1992) 587–597.
- [39] R.D. Hays, et al., Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items, *Qual. Life Res.* 18 (7) (2009) 873–880.
- [40] G.G. Fillenbaum, *Multidimensional Functional Assessment of Older Adults: The Duke Older Americans Resources and Services Procedures*, Erlbaum Assoc., Inc., Hillsdale, 1988.
- [41] B.W. Smith, et al., The brief resilience scale: assessing the ability to bounce back, *Int. J. Behav. Med.* 15 (3) (2008) 194–200.
- [42] D. Ferriman, J.D. Gallwey, Clinical assessment of body hair growth in women, *J. Clin. Endocrinol. Metab.* 21 (1961) 1440–1447.
- [43] R.S. Swerdloff, et al., Long-term pharmacokinetics of transdermal testosterone gel in hypogonadal men, *J. Clin. Endocrinol. Metab.* 85 (2000) 4500–4510.
- [44] M. Padero, S. Bhasin, T. Friedman, Androgen supplementation in older women: too much hype, not enough data, *J. Am. Geriatr. Soc.* 50 (2002) 1131–1140.
- [45] F.M. Gloth, J.D. Tobin, Vitamin D deficiency in older people, *J. Am. Geriatr. Soc.* 43 (1995) 822–828.