STRATEGIES TO REDUCE INJURIES AND DEVELOP CONFIDENCE IN ELDERS (STRIDE)

RANDOMIZED TRIAL OF A MULTIFACTORIAL FALL INJURY PREVENTION STRATEGY

STATISTICAL ANALYSIS PLAN

Version 2.0

November 25, 2019

The SAP was prepared by the blinded trial statisticians:
Peter Peduzzi, PhD
Denise Esserman, PhD
Michael Miller, PhD
Thomas Travison, PhD

Revisions to the SAP were done before unblinding of the data.

Modification to the SAP on 2-10-2020 for the Analysis of the Secondary Outcome All Falls

For the secondary outcome, all falls, the dates of events were not recorded. Thus, the analytic plan for first fall was modified from a time-to-event analysis to a practice-level Poisson model using time from baseline to the midpoint of the interview window in which the first fall was recorded as an offset in the model. [Note, the practice-level model was the default model when issues were encountered with patient-level models.] The lack of actual event times also precluded the calculation of cumulative incidence rates, which were replaced by event rates per person year of follow up. The analysis of recurrent falls was analyzed using the same method as the other recurrent event outcomes, i.e., a practice-level Poisson model with an offset for total follow-up time.

Modification to the SAP on 8-1-2020 for the Analysis of the Secondary Outcome All Falls

For the secondary outcome, all falls, the dates of events were not recorded and the exact counts of the number of events were also not recorded. Thus, a proper analysis of time to first fall and recurrent falls could not be done.

TABLE OF CONTENTS

STUDY SUMMARY (revised 7/31/2018)

Title	RANDOMIZED TRIAL OF A MULTIFACTORIAL FALL INJURY PREVENTION STRATEGY
Study Design	The design is a cluster randomized, parallel group superiority trial with practices stratified by healthcare system and patients nested within practices. The unit of randomization is the practice.
Study Duration	6 years
Trial Sites	10 trial sites: The Partners' Health Care System; Essentia; Hopkins Health Care System; HealthCare Partners; Reliant Health Care System; Mount Sinai Health Care System; University of Pittsburgh Health Care System; University of Texas Medical Branch Health Care System; University of Iowa Health Care System; University of Michigan Health Care System.
Objective	Conduct a cluster-randomized trial to determine the effectiveness of an evidence-based, patient-centered multifactorial fall injury prevention strategy.
Number of Subjects	The original target sample size was 6,000 participants enrolled from 86 practices to provide 90% power to detect a 20% reduction in the rate of the primary outcome with intervention relative to control. Later, the duration of the trial was extended to a total of 40 months (20 months of recruitment and an additional 20 months of follow-up), which reduced the target sample size to 5,322 participants. Recruitment ended after 20 months on March 31, 2017, with a total of 5,451 participants enrolled.
Main Inclusion Criteria	Community-living persons, 70 years or older, who are at increased risk for serious fall injuries.
Intervention	An evidence-based patient-centered intervention that combines elements of a multifactorial, risk factor-based, standardly-tailored fall prevention strategy developed at Yale, practice guidelines offered by the CDC's "STEADI" toolbox and the joint American Geriatrics Society/British Geriatrics Society guidelines, and ACOVE practice change approach.
Duration of Intervention	A minimum of 20 months and a maximum of 40 months.
Primary Outcome	The primary outcome is serious fall injuries, operationalized as a fall resulting in: (1) (fracture other than thoracic/lumbar vertebral; joint dislocation; or cut requiring closure) AND any medical attention; OR (2) (head injury; sprain or strain; bruising or swelling; or other) requiring hospitalization.
Primary Analysis	The risk of any serious fall injury (i.e., time to first event) will be analyzed using a survival model that incorporates competing risks (due to death) and clustering. In this analysis, participants who are lost to follow-up without a prior serious fall-related injury will be censored at their date last seen. In a secondary analysis, we will adjust for the pre-specified set of baseline covariates to examine their influence on the intervention effect.

Secondary Outcomes	Number of falls, number of all fall injuries, and measures of well-being.
Adaptive Components	**Adaptive components of the trial** include: 1) monitoring the accrual rate to determine whether the study eligibility criteria need to be reconsidered if recruitment is lower than expected, taking into account that any changes could affect the inferences; 2) monitoring the potential for ascertainment bias because of interactions between the FCMs and study participants and changing the primary outcome definition if necessary; 3) monitoring the primary outcome rate to determine whether the outcome needs to be adapted, e.g., from time to first serious fall-related injury to time to all recurrent serious fall-related injuries if the former rate is too low, affecting the power of the study; 4) interim monitoring for efficacy or futility, if necessary; and 5) refining the analytic methods based on the validity of the assumptions; such an adaptation will be done blinded to treatment, e.g., if the death rate is low, competing risks could be considered as a secondary, rather than a primary, analysis.
Interim Analysis	Interim monitoring will focus on patient accrual, baseline comparability of treatment groups, protocol adherence, data completeness and quality, accrual of fall events, safety, and efficacy or futility.

PROTOCOL VERSION HISTORY (revised 7/31/2018)

Version	Date	Changes
2.7	7/31/2018	1. Extend the study from a minimum of 20 and maximum of 40 months follow-up to a minimum of 24 and maximum of 44 months of follow-up. Final end date of the study, including the analytic phase is now extended to April 30, 2020.
2.6	6/1/2018	1. The definition of the primary outcome has been modified to reflect the recommendations of the Ad Hoc Expert Panel established by the NIA and NIA's concurrence with this recommendation.
2.5	12/13/2017	1. Duration of the study changed from 36 months (18 months of recruitment and a minimum 18 months of follow-up) to 40 months (20 months of recruitment and a minimum of 20 months of follow-up). 2. Target sample size adjusted to n=5,322 for a 40-month study instead of n=6000 for a 36-month study. 3. Interim monitoring was changed from efficacy and futility to efficacy or futility, if necessary.
2.4	6/27/2017	1. Clarification of the Adjudication process with CMS
2.3	1/28/2016	1. Changed the lower age limit for eligibility from 75 to as low as 70, if needed based on recruitment.
2.2	6/19/2015	1. Eliminate the postal questionnaire to ascertain fall/fall injury data. Instead, obtain these data through telephone calls from the RAC to all participants. (Note: all STRIDE participants will continue to use the fall calendars as a memory aid). 2. Change the timing of follow-up calls to patients to inquire about falls/fall injuries from every 3 months to every 4 months. 3. To comply with Pennsylvania state law, the list of people who are eligible to serve as proxy respondents for people who are unable to provide informed consent due to cognitive or hearing impairment has been adjusted.
2.1	3/25/2015	Original

1. BACKGROUND

Approximately one in three older Americans falls each year and 20-30% of those who fall suffer moderate to severe injuries such as lacerations, hip fractures, or head trauma. The problem is important particularly among those 75 years and older when the incidence of falls rises dramatically. Among older adults, falls are the leading cause of both fatal and nonfatal injuries. In 2010, 2.3 million nonfatal fall injuries were treated in emergency departments and more than 662,000 of these patients were hospitalized. These numbers will rise with the aging of the baby boomers. In addition, many who do not sustain injuries develop fear of falling, which may result in self-limiting their activities, leading to reduced mobility and loss of physical fitness, further increasing their risk of falling.

Despite decades of research that has demonstrated conclusively that many falls in the elderly can be prevented, quality of care for falls remains low. Fewer than half of those who fall each year discuss their falls or fall prevention with a health care provider. Moreover, only a third of elderly patients in primary care practice are screened for falls and the quality of care for those who are at risk for falling has not improved in the past decade.

Accordingly, this project funded jointly by the National Institute on Aging and the Patient Centered Outcomes Research Institute (PCORI) is a multi-site cluster randomized clinical trial (RCT) of an evidence-based, multifactorial individually-tailored intervention to reduce the risk of serious fall injuries among non-institutionalized older persons.

2. AIMS

1. To conduct in partnership with patients and stakeholders a cluster randomized trial of the intervention in a sample of 6,000 participants from 80+ practices in 10 health care systems that reflect geographic and sociodemographic diversity. The trial will include features of an adaptive design, to facilitate “learning”, including a pre-specified, mid-term analysis.
2. To identify and evaluate jointly with patients and stakeholders the obstacles and facilitators of the post-trial scalability, sustainability, and dissemination of the fall injury prevention strategy.

3. STUDY DESIGN

STRIDE is designed as a cluster randomized, parallel group superiority trial with practices stratified by healthcare system and patients nested within practices. The unit of randomization is the practice to avoid the potential for contamination of controls, which could occur more easily if randomization was at the level of the participant or physician. The primary outcome is time to first serious fall injury assessed at the patient level. The original target sample size was 6000 participants from a target of 86 practices to detect a 20% reduction in serious fall injuries in the intervention group versus control group with 90% power. The target sample size was revised to 5,322 in December 13, 2017; recruitment ended after 20 months on March 31, 2017, with a total of 5,451 participants enrolled.

STRIDE was designed with the following adaptive features:

1. Monitoring the accrual rate to determine whether the eligibility criteria should be altered if enrollment is lower than expected (e.g., by lowering the age entry criterion);
2. Monitoring the potential for ascertainment bias because of interactions between the FCMs and study participants and changing the primary outcome definition, if necessary;
3. Monitoring the primary outcome rate to determine whether the outcome needs to be adapted, e.g., from first serious fall injury to all serious fall injuries if the former rate is too low, affecting the power of the study;
4. Interim monitoring for efficacy and futility, if necessary; and
5. Refining the analytic methods based on the validity of the assumptions; such an adaptation will be done blinded to treatment, e.g., if the death rate is low, competing risks could be considered as a secondary, rather than a primary, analysis.

4. OUTCOMES

The primary and secondary outcome measures are summarized in Table 1.

Table 1. Primary and Secondary Outcome Measures, assessed centrally by Yale RAC

Domain	Measure (1˚ & 2˚)	Source, Frequency, and Sample
Fall-related	Serious fall injuries (1˚)	Telephone interview every 4 months, supplemented by administrative claims/encounter data (including data from clinical trial sites and Medicare) for date and type of injury; medical record adjudication for discrepancies with fall-injury claims
	All fall injuries (2˚)	Telephone interview every 4 months: serious fall injuries plus other injuries that may not come to medical attention
	Original operationalized definition of primary outcome (2˚)	Telephone interview every 4 months, supplemented by administrative claims/encounter data (including data from clinical trial sites and Medicare) for date and type of injury
	All falls (2˚)	Telephone interview every 4 months
Well-being	Concern about falling (2˚)	Modified FES at baseline, 12 m and 24 m (telephone) in 720 participants, representing a random sample of participants enrolled in first 12 months
	Physical function and Disability (2˚)	LL-FDI via CAT at baseline, 12 m and 24 m (telephone) in 720 participants
	Anxiety/Depressive symptoms (2˚)	PROMIS scales measured at baseline, 12 m and 24 m (telephone) in 720 participants

CAT: Computer Adaptive Testing; FES: Fall Efficacy Scale; LL-FDI: Late-Life Function and Disability Instrument; PROMIS: Patient Reported Outcomes Measurement Information System

4.1 Primary Outcome

The primary outcome is serious fall injuries leading to medical attention, including non-vertebral fractures, joint dislocation, head injury, lacerations, and other major sequelae (e.g., rhabdomyolysis, internal injuries, hypothermia). A fall is defined using the traditional definition of a person unintentionally coming to rest on the ground or other lower level not as a result of a major intrinsic event (e.g. myocardial infarction or stroke) or an overwhelming external hazard (e.g. hit by a vehicle). The following will not be included as falls: a controlled or intentional movement to a chair or bed; a "near fall" in which the participant caught himself or herself before hitting the floor, ground, or object; and being knocked down by a substantial external force, like a moving vehicle. Fall injuries from height > 3 feet will not be included.

The definition of the primary outcome was modified on 6/1/2018 to reflect the recommendations of the Ad Hoc Expert Panel established by the NIA and NIA's concurrence with this recommendation. The definition of a serious fall injury was adapted to:

A fall resulting in:
(1) fracture other than thoracic/lumbar vertebral; joint dislocation; or cut requiring closure AND any medical attention;
OR
(2) head injury; sprain or strain; bruising or swelling; or other requiring hospitalization.

The original definition only required medical attention for a head injury; sprain or strain; bruising or swelling; or other – not a hospitalization.

4.2 Secondary Outcomes

The secondary outcomes include all fall injuries and all falls (regardless of injury), as well as indicators of well-being. All fall injuries include serious fall injuries as well as less severe falls that result in bruises, cuts, persistent pain, and restricted activities, but not necessarily medical attention. The four indicators of well-being are: fall efficacy, physical function/disability, anxiety, and depressive symptoms administered by brief instruments over the phone. The indicators of well-being will be measured at 12 and 24 months in a random subsample of 720 participants. The random subsample will be selected from year 1 enrollees to permit 24 month assessments within the trial duration of 3 years.

4.3 Tertiary Outcomes

The tertiary outcomes include hospitalizations and long-term nursing home admissions.

5. RANDOMIZATION

5.1 Method of Randomization

All participating clinical practices that met trial eligibility criteria were randomized at one time using covariate-constrained randomization with stratification by clinical site (healthcare system) to control for potential site differences. Covariate-constrained randomization was used to balance practice characteristics within and across the clinical sites (strata). The balancing

covariates were practice size, geography (urban vs. rural), and race/ethnicity (primary identification nonwhite vs. white).

A published SAS macro was used to conduct the covariate-constrained randomization (Chaudhary and Moulton, 2006; Greene, 2017). The macro generates a set of randomizations that satisfy balance on practice size, geography and race/ethnicity both within and across strata. From this set of valid assignments, one was selected at random. Only the trial biostatisticians participated in the generation of the randomization. No one else from the trial was involved in the process. To minimize the risk of selection bias, practice site names were masked during the process. The practice site randomization assignments to treatment groups were released to the clinical sites only after careful vetting of the entire randomization process by the trial biostatisticians.

Practices will not be added after trial initiation. All available practices that met our eligibility criteria have been carefully vetted and have committed to the trial. The criteria included: sufficient number of age eligible patients (≥ 400), access to community resources (specifically, exercise programs), access to electronic health records (EHR) and availability of practice characteristics. Thus, additional practices are not readily available. The participating practices have more than enough age eligible patients (> 80,000) to meet the original target recruitment goal of 6,000 patients, assuming a conservative 10% enrollment rate, which is reasonable based on the pilot study data. If a randomized practice drops out of the trial, we expect to ascertain study outcomes in individual participants since follow-up is being conducted centrally.

5.3 Allocation Concealment

An important consideration, particularly in cluster trials, is allocation concealment to control for selection bias. Potential approaches include 1) randomizing after enrollment, 2) blinding the recruiters, 3) standardizing the enrollment process with adequate training of screeners and recruiters, and 4) covariate adjustment in the analysis. Because enrollment must occur after practices have been randomized in STRIDE, the first approach is not feasible. However, the possible effects of selection bias will be mitigated through approaches 2 and 3, along with covariate adjustment in the analysis, if necessary.

6. SAMPLE SIZE

6.1 Preliminary Data

Sample Size estimates were informed by preliminary data from MOBILIZE Boston, a longitudinal study of 765 community-living seniors, with monthly ascertainment of serious fall injuries. Although based solely on self-report, the operational definition of serious fall injury was otherwise the same as in the proposed trial. Analyses were restricted to 135 participants meeting STRIDE entry criteria, with a median follow-up of 2.8 years. The annual rate of first serious fall injury was 18% (95%CI: 14%-24%); a preliminary estimate of the overall serious fall injury rate (including recurrent events) was comparable, at 21% (95%CI: 16%-24%). For the proposed trial, we project an annual outcome rate in the control group of 14% to 18%, since fall injuries will be assessed every three months rather than monthly, and our protocol requires that serious fall injuries be confirmed by data from claims or medical records.

Data from the Lifestyle Interventions and Independence for Elders (LIFE) Study and Assessing Care of Vulnerable Elders Practice Redesign for Improved Medical Care for Elders (ACOVEprime) were also examined in order to inform the outcome rate, but LIFE had a lower risk population and ACOVEprime primarily used Medicare claims data to ascertain fall rates. Data from MOBILZE Boston and LIFE were used to help inform the loss and death rates. The annual loss rate in the absence of serious fall injury was 4% in MOBILIZE Boston, but this study did not have access to claims or EHR data. The rate in LIFE was less than 2% in 728 patients aged 75 years and older. The annual death rates in the absence of serious fall injury were less than 2% for both MOBILIZE Boston and the LIFE Study. Because the MOBILIZE Boston rate was based on only a few number of deaths and the LIFE Study enrolled a lower risk population because the intervention included an exercise component, we assumed a higher death rate for STRIDE.

6.2 Sample Size Determination for the Primary Outcome

Sample size was determined for a clustered design for time to first serious fall injury (the primary outcome) based on the logrank statistic in the presence of competing risk due to death. Table 2 presents sample size estimated using PASS version 12 (Kaysville, Utah) under the following assumptions:

1. type I error = 5% (2-sided) and 90% power;
2. a trial duration of 3 years with accrual periods of 1, 1.5 and 2 years (assuming uniform accrual);
3. equal allocation to intervention and control groups;
4. no adjustment for non-adherence to intervention (accounted for with conservative treatment effect);
5. all patients followed to end of the trial (max = 3 years);
6. a 20% reduction in serious fall injuries with intervention relative to control, assuming constant and proportional hazards. [This reduction, which would have considerable clinical and public health impact, is intermediate between that found in Yale FICSIT (31% reduction in all falls) and Connecticut Collaborative Falls Program (CCFP) (9% reduction in serious fall injuries), two prior multifactorial fall prevention studies. Because many aspects of our intervention will be implemented at the level of the patient (as in FICSIT) rather than geographic region (as in CCFP), our effect size should be closer to that of FICSIT than CCFP, making the sample size estimates given here conservative.]
7. 7% annual death rate without experiencing a serious fall injury (i.e., competing risk);
8. 3% annual loss to follow-up rate in the absence of serious fall injury or death (expected to be low due to the use of multiple sources, including claims and medical records data;
9. 3% inflation for the proposed interim monitoring for efficacy and futility; and
10. 53% inflation for the design effect of clustering, assuming an average of 70 participants enrolled per practice from 86 practices and an ICC of 0.0076 estimated from an analysis of serious fall injuries in the LIFE Study.

Based on Table 2, a target sample size of 6000 participants was selected to be accrued from 86 practices, each enrolling an average of 70 participants over a 1.5-year recruitment period. This sample size provided 90% power to detect a 20% reduction in the hazard rate with intervention

relative to control for an annual first serious fall injury rate of 0.14 in the control group. The target number of serious fall injuries to detect this reduction is 845 events.

On December 13, 2017, a protocol amendment was approved extending the duration of the study from 36 months (18 months of recruitment and a minimum 18 months of follow-up) to 40 months (20 months of recruitment and a minimum of 20 months of follow-up). The target sample size was adjusted to 5,322 participants for a 40-month study instead of 6000 participants for a 36-month study. Recruitment ended after 20 months on March 31, 2017, with a total of 5,451 participants enrolled.

Table 2. Sample Size Estimates

Accrual Period (Years)	Annual Event Rate in the Control Group	Total Sample Size
1	0.14	5532
	0.16	4906
	0.18	4418
1.5	0.14	6030
	0.16	5332
	0.18	4792
2	0.14	6666
	0.16	5888
	0.18	5284

6.3 Power for Secondary Outcomes

Power for all fall injuries and all falls (regardless of injury) was calculated using the same assumptions as above, but with a 1% Type I error to control for multiplicity and no adjustment for interim analysis. Annual rates of first fall and first fall injury from MOBILIZE Boston are 0.92 (95%CI: 0.75, 1.10) and 0.57 (95%CI: 0.46, 0.69), respectively. Using conservative rates of 0.70 and 0.40, and based on the targeted sample size of 6000, power is > 99% to detect a 20% reduction with intervention relative to control for both of these secondary outcomes. For the indicators of well-being, an unadjusted sample size of 524 participants is required to detect a standardized effect size of 0.3 between intervention and control at 1 and 2 years with 80% power and 1% type I error. The sample size adjusted for missing outcomes due to death (10%/year) and lost to follow-up (5%/year with or without a serious fall injury), and for clustering (DE = 6% assuming ICC = 0.0076 with 86 practices each enrolling 8 participants) is 720 (360 per group), a 12% subsample of the 6000 participants to be accrued during year 1 of the trial.

6.4 Power for Tertiary Outcomes

There were no data to inform power calculations for the two tertiary outcomes, hospitalizations and long-term nursing home admissions.

7. INTERIM MONITORING PLAN

7.1 Overview

Interim monitoring will focus on patient accrual, baseline comparability of treatment groups, protocol adherence, loss to follow-up, data completeness and quality, sample size assumptions, accrual of fall events, safety, efficacy, and futility. The summary of the types of tables, listings and figures (TLFs) generated for semi-annual open and closed DSMB reports is presented in the Appendix A. The final Interim Monitoring Plan (dated May 17, 2017) is given in Appendix B and was approved by the DSMB at its May 31, 2017 meeting.

7.2 Trial Adaptation Guidelines

Adaptive design features proposed in the protocol include: 1) monitoring the accrual rate to determine whether the study eligibility criteria need to be reconsidered if recruitment is lower than expected, taking into account that any changes could affect the inferences; 2) monitoring the potential for ascertainment bias because of interactions between the FCMs and study participants and changing the primary outcome definition if necessary; 3) monitoring the primary outcome rate to determine whether the outcome needs to be adapted, e.g., from time to first serious fall-related injury to time to all recurrent serious fall-related injuries if the former rate is too low, affecting the power of the study; 4) interim monitoring for efficacy or futility, if necessary; and 5) refining the analytic methods based on the validity of the assumptions; such an adaptation will be done blinded to treatment, e.g., if the death rate is low, competing risks could be considered as a secondary, rather than a primary, analysis.

Any adaptations will be presented to the DSMB for their review and approval prior to implementation. The DCC will designate an independent (unblinded) statistician for reporting to the DSMB in closed session. The independent statistician will not be involved in any operational issues in the trial once it is initiated. Only the independent statistician, along with the statistical programmer, will have access to the link between coded (A vs. B) and actual treatment assignment (intervention vs. control). All other DCC and trial personnel will remain blinded to the actual treatment assignment.

Any adaptation of eligibility criteria because of inadequate recruitment will be based on analyses of aggregate baseline screening data to determine the reasons for study ineligibility. Since these data are obtained at baseline and prior to the initiation of treatment, the analyses can be considered to be unbiased. It is expected that the study leadership will be involved in this type of adaptation.

Adaptation based on a lower than expected primary outcome rate will be done using aggregate data and presented to the DSMB in closed session by the independent statistician. No other DCC or trial personnel will be privy to these data because knowledge of the outcome rate could affect any operational decisions for the trial.

Any interim analyses for efficacy and futility will be presented in closed session to the DSMB by the independent statistician. Treatments will be designated as A vs. B. The independent statistician will have access to the actual treatment assignment during the closed session, if requested by the DSMB. No one else associated with the trial will be privy to these data.

Refining the Statistical Analysis (SAP) plan will be done blinded to treatment by the DCC blinded trial statisticians. If the outcome is changed the DCC statisticians will need to be aware of the change, in addition to the independent statistician, in order to be able to appropriately revise the SAP.

7.3 Monitoring Inter-Practice Variation

Because of the concern raised by the DSMB at its initial meeting about inter-practice variation and its potential effect on the power of the trial, we monitored the effect of clustering using a variance inflation factor (VIF) approach. That is, the inflation in variance for the treatment effect for the model with clustering vs. the model without clustering. This approach was used because of the complexity of the analytic model, a clustered survival model with competing risk due to death, and the lack of a closed form ICC for this model. The VIF was determined by the unblinded trial statisticians and presented in the closed DSMB session.

7.4 Monitoring of Efficacy and Futility

Per RFA, a “preliminary statistical analysis of the effects of the preventive strategy on the rate of serious fall injuries must be reported in March 2017.” At this time, recruitment will have ended and about 55% of the expected number of primary events will have been accrued, sufficient for an interim look. The expected number of events is based on the ratio of average follow-up time at the interim look (1.25 years) relative to the average total follow-up time for the study (2.25 years), assuming recruitment begins on June 1, 2015. Thus, we propose consideration of stopping the trial at this time only for compelling evidence of efficacy (~$p<0.008$) or a trend in the wrong direction (i.e., futility) taking into account trends in secondary outcomes. Future looks will be left to the purview of the DSMB based on emerging trends in the data; however, by March 2017, there will be less than one year left in the trial and it may be impractical to conduct another interim look before it ends in February 2018. Interim monitoring boundaries will be established using an alpha spending approach using EAST version 6.13 (Cytel) software (Jennison and Turnbull, 2000; Proschan, Lan and Wittes, 2006). A prototype figure is given in the Appendix.

In addition to formal sequential monitoring boundaries, estimates of conditional power will be provided to assess the likelihood of reaching a significant conclusion at the end of the trial given the observed data (Lan and Wittes, 1988).

On December 13, 2017, a protocol amendment was approved changing the interim monitoring (with DSMB approval) from efficacy and futility to efficacy or futility, if necessary.

7.5 Monitoring of Safety

Details of the safety and adverse event monitoring plan were provided in the study protocol. Because STRIDE is a low risk trial and an experimental intervention is not being used, only serious adverse events related to hospitalization and death will be collected. Adverse events will not be collected. The vast majority of SAE ascertainment will occur from postal questionnaire and telephone follow-up and/or in aggregate via biannual data downloads from the EHR and claims data. Unanticipated problems will also be reported directly to the Medical Safety Monitor from the sites.

The distribution of hospitalizations, deaths and types of unanticipated problems will be summarized overall and by treatment. Death rates will be calculated by the method of Kaplan-Meier. A listing of the safety tables and figures to be presented to the DSMB are summarized in the Appendix.

8 ANALYTIC PLAN

8.1 Overview

The analysis of the primary and secondary outcomes will be according to the principle of intent-to-treat, i.e., practices/participants will be analyzed according to their original treatment assignment regardless of adherence to protocol. All analyses will account for the cluster design with the participant as the unit of analysis. SAS 9.4 and R 3.6.1 software will be used for all analyses.

8.2 Comparability of Treatment Groups

Comparability of treatment groups will be assessed by comparing the distribution of baseline characteristics in the two groups using appropriate graphical procedures, summary statistics and multivariate methods. The randomization is designed to produce balance on important covariates at the practice level (unit of randomization) but not necessarily at the patient level. Therefore, in a secondary analysis, the following pre-specified set of baseline covariates will be selected for adjustment to determine their influence on the treatment comparisons: age, race/ethnicity, education, number of chronic conditions, and number of positive screening items (out of three) for serious fall injuries (Senn, 1989).

Note: Sections 8.3 – 8.5 were revised to provide more analytic detail than presented in version 1. 0 of the SAP, including more recent statistical advances. The revised plan is consistent with the original analytic plan. The trial protocol has been revised accordingly.

8.3 Analysis of Primary Outcome: Adjudicated Serious Fall Injury

8.3.1 Primary Analysis of Primary Outcome: Adjudicated Confirmed First Event

Analysis	Primary analysis of primary outcome: adjudicated confirmed first serious fall injury
Analysis population	All enrolled participants
Endpoint	Time to first adjudicated confirmed serious fall injury
Unit of analysis	Participant
Method of analysis*	Multistate survival model that incorporates competing risks due to death and clustering (Putter et al., 2007; Lee et al., 1992; R

	Package 'survival': https://cran.r-project.org/web/packages/survival/survival.pdf)
Calculation of follow-up time	Time from date of enrollment to the earliest of: date of first confirmed serious fall injury, date of last completed patient or proxy interview, or date of death.
Adjustment covariates	Cluster-level covariates: healthcare system; constrained-covariate randomization balancing factors: practice size (tertiles), geography (urban vs. rural), and race/ethnicity (majority primary identification: nonwhite vs. white)
Type I error	5% (2-sided)
Control of type I error	None
Treatment effect estimate	Hazard ratio with 95% confidence limits

* If the multistate model does not converge, we will fit the Zhou et al. (2011) model. In the event there are practices with no events that create convergence issues with fitting patient-level models for the primary outcome [i.e., multistate and Zhou et al. (2011)], the primary analysis will default to a practice-level analysis as described below.

Sensitivity analysis: Adjustment for pre-specified set of patient-level covariates (in addition to cluster-level covariates): age, sex, race/ethnicity, education, number of chronic conditions, and number of positive screening items for serious fall injuries.

Subgroup analysis: The effect of intervention will be evaluated in the following prespecified subgroups of participants using appropriate tests of homogeneity (e.g., interaction): age (70-79, 80+), sex (male, female), fear of falling alone (yes, no), multimorbidity (0-1 chronic conditions, 2+ chronic conditions), and hip fracture or other fracture since age 50 (yes/no)]. The Hochberg (1988) procedure will be used to control for multiplicity for the tests of interaction using an overall type I error of 5% (2-sided).

Cumulative incidence rates: The cumulative incidence of first serious fall injury will be estimated using non-parametric maximum likelihood methods (Aalen-Johansen estimator; Putter et al, 2006) and will be used to estimate freedom from falling over the entire follow-up period.

Supplementary Practice-level analysis: This analysis will be based on a Poisson regression model using total practice-level follow-up time as an offset, adjusted for healthcare system and the cluster-level covariates [healthcare system; constrained-covariate randomization balancing factors: practice size (tertiles), geography (urban vs. rural), and race/ethnicity (primary identification: nonwhite vs. white)]. [We note that the practice-level analysis is a planned supplementary analysis and will default to the primary analysis only if the multistate and Zhou models do not converge.]

8.3.2 Supportive Analysis of the Primary Outcome: Adjudicated Confirmed Recurrent Events

Analysis	Supportive analysis of primary outcome: all adjudicated confirmed serious fall injuries (recurrent events)
Analysis population	All enrolled participants
Endpoint	Times to recurrent adjudicated confirmed serious fall injuries
Unit of analysis	Participant
Method of analysis*	Joint frailty model with recurrent events and death as a semi-competing risk (Jung et al., 2018).
Calculation of follow-up time	Time from date of enrollment to the earlier of date of the last completed patient or proxy interview, or date of death.
Adjustment covariates	Cluster-level covariates: healthcare system; constrained-covariate randomization cluster-level balancing factors: practice size (tertiles), geography (urban vs. rural), and race/ethnicity (majority primary identification: nonwhite vs. white)
Type I error	5% (2-sided)
Control of type I error	None
Treatment effect estimate	Hazard ratio with 95% confidence limits

*If the Jung model does not converge, we will fit a practice-level analysis as described above.

8.4 Analysis of Secondary Outcomes

The secondary outcomes are self-reported fall injuries, self-reported falls and indicators of well-being (fall efficacy, physical function, anxiety and depressive symptoms). Also included as a secondary outcome is unadjudicated patient-reported serious fall injuries under the original definition. To provide some control for multiplicity, we will test the secondary outcomes using a significance level of 1% (2-sided) and report 99% confidence limits.

8.4.1 All Fall Injuries and All Falls

These endpoints will be based on patient-reported events and analyzed as both time to first event and recurrent events using the methodology outlined for the primary outcome. Patient-reported serious fall injuries under the original definition will be analyzed similarly.

8.4.2 Indicators of Well-Being

Indicators of well-being include fall efficacy (measured by FES), physical function (measured by LL-FDI), anxiety and depressive symptoms (measured by Promise scales for anxiety and depression) and were measured at 12 and 24 months in a random subsample of 714 participants. Each outcome will be analyzed separately as described below.

Analysis	Indicators of well-being
Analysis population	Random subsample of 714 participants
Endpoint	Scale values at 12 and 24 months of follow-up
Unit of analysis	Participant
Method of analysis	Linear mixed model assuming missing at random (MAR)
Handling of missing data	MAR assumption
Adjustment covariates	Baseline indicator of well-being; Factors predictive of missingness: these will be determined at the time of analysis
Type I error	1% (2-sided)
Treatment effect estimate	Marginal least square mean (LSM) with 99% confidence limits if there is no treatment by time interaction ($p<0.10$); otherwise, LSM will be summarized for each treatment and timepoint

Sensitivity analyses will be conducted to investigate the MAR assumption, such as methods that model jointly the missingness and outcome distributions (National Research Council, 2010).

8.5 Analysis of Tertiary Outcomes

Hospitalizations will be analyzed using a Poisson regression model accounting for practice-level clustering with follow-up time as an offset. Time to first long-term nursing home admission will be analyzed using the multistate model accounting for clustering (as described for primary outcome). Death will be analyzed using the marginal Cox model (Lee et al, 1992). No control for multiplicity will be done for these analyses.

8.6 Analysis of Safety

Safety analyses will involve tabulating the occurrence of serious adverse events (deaths and hospitalizations) and unanticipated problems between the two groups. Hospitalizations and death will be analyzed as described in section 8.5. A p-value of 0.05 (2-sided) will be used for the safety analyses. No control for multiplicity will be done for the safety analyses.

REFERENCES

Chaudhary MA, Moulton LH (2006). A SAS macro for constrained randomization of group-randomized designs, Computer Methods and Programs in Biomedicine. 83:205-210.

East: Software for Advanced Clinical Trial Designs, Simulation and Monitoring. Version 6.1. Cambridge, Massachusetts: Cytel Statistical Software and Services, 2013.

Greene EJ (2017). A SAS macro for covariate-constrained randomization of general cluster-randomized and unstratified designs. Journal of Statistical Software. April 2017, Volume 77, Code Snippet 1. doi: 10.18637/jss.v077.c01

Hochberg Y (1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 75:800-802.

Jennison C, Turnbull BW. Group Sequential Tests with Applications to Clinical Trials. Boca Raton, Florida: Chapman and Hall/CRC; 2000.

Jung T, Peduzzi P, Allore HG, Kyriakides TC, Esserman D (2018). A joint model for recurrent events and a semi-competing risk in the presence of multi-level clustering. Statistical Methods in Medical Research 2018 Jul 31:962280218790107. doi: 10.1177/0962280218790107. [Epub ahead of print] PMID:30062911

Lan KKG, Wittes J (1988). The B-value: A tool for monitoring data. Biometrics. 44:579-5858.

Lee EW, Wei LJ, Amato DA (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In Survival Analysis: State of the Art, JP Klein, PK Goel (eds.). pp 237-247, Klawer Academic Publishers.

National Research Council (2010). The prevention and treatment of missing data in clinical trials. National Academies Press.

Proschan MA, Lan KKG, Wittes JT (2006). Statistical Monitoring of Clinical Trials: A Unified Approach. New York: Springer.

Putter H, Fiocco M, Geskus R (2007). Tutorial in biostatistics: competing risks and multi-state models. Statistics in Medicine. 26(11):2389-2430.

Senn S (1989). Covariate imbalance and random allocation in clinical trials. Statistics in Medicine. 8:467-475.

Zhou B, Fine J, Latouche A, Labopin M (2011). Competing risks regression for clustered data. Biostatistics.13(3):371-383.

APPENDIX A: DSMB TABLES, LISTINGS AND FIGURES (TLFs)

The general types of table, listings and figures (TLFs) generated for the Open DSMB Report included:

1. Intake graph showing cumulative observed and expected intake
2. Consort diagram showing flow of screening and recruitment data
3. Reasons for exclusion overall and by site
4. Entry characteristics of enrolled participants overall and by site
5. Completeness of follow-up overall and by site: person years of follow-up, losses and withdrawals
6. Fidelity of treatment
7. Frequency of serious adverse events (SAEs) overall and by site: mortality and hospitalizations
8. Listing of SAEs
9. Protocol Deviations

The closed DSMB Report mirrored the Open Report and presented data by randomized treatment group designated as A and B. Additional closed report tables included information about power and primary and secondary outcome data.

APPENDIX B

STRIDE Trial
Interim Monitoring Plan
May 17, 2017

Introduction

The STRIDE study was designed as an adaptive trial. Originally, the adaptive components of the trial included: 1) monitoring the accrual rate to determine whether the study eligibility criteria need to be reconsidered; 2) monitoring the primary outcome rate to determine whether the outcome needs to be adapted, e.g., from time to first serious fall-related injury to time to all recurrent serious fall-related injuries if the former rate were too low; and 3) interim monitoring for efficacy and futility. The original proposal envisioned an interim look in March 2017, but that date was considered unrealistic because of the trial's late start and the accrual of insufficient numbers of adjudicated events by that date. Accordingly, the sponsors and the DSMB left the exact timing and content of the interim monitoring to the discretion of the investigators. In its November 2016 meeting, the trial's DSMB directed the study team to present an interim monitoring plan for consideration of the DSMB in its May 2017 meeting.

In response to the DMSB, the study team has prepared this document describing the Interim Monitoring Plan (**IMP**) for the NIA/PCORI-funded Strategies to Reduce Injuries and Develop Confidence in Elders (**STRIDE**) Trial. The IMP is specifically intended to assist the STRIDE Trial DSMB in forming recommendations concerning ongoing trial progress and potential design changes. Specifically, it focuses on three critical decision points for review over the next year: (i) change in the primary outcome definition; (ii) extension of follow-up or adaptation of the primary endpoint to a more inclusive "all events" implementation to maintain study power; and (iii) cessation or continuation of trial procedures as informed by a futility analysis.

Organization of the IMP

The **IMP** is divided into three stages to address the following issues:

- Stage 1. Evaluation of whether the primary outcome definition should be changed to address the potential for ascertainment bias resulting from differential rates of referral for medical care in the setting of a fall
- Stage 2. Evaluation of event accumulation and decision to extend follow-up or adapt the primary outcome to recurrent events
  - 2a. Monitoring the cluster design effect
  - 2b. Monitoring the accrual of adjudicated primary study outcomes in the control group
  - 2c. Considerations to extend follow-up to maintain study power for first serious fall injury (SFI) or adapt the primary outcome from first SFI to all SFI
- Stage 3. Timing and nature of the planned interim analysis.

These issues are complex and need to be addressed before taking a formal "look" at the intervention effect for efficacy/futility and must be done blinded to outcome data by intervention group. All analyses to implement the IMP will be performed by the unblinded trial statisticians. A brief description of each stage follows.

Stage 1 is to determine whether the primary outcome definition needs to be changed to address concerns about potential ascertainment bias because of interactions between the fall care managers and the study participants. Such bias has the potential to dilute the intervention effect affecting study power.

Stage 2 addresses the accumulation of primary outcomes and study power. Specifically, whether the generation of primary study outcomes in the control group is on target to maintain the power of the trial to detect the hypothesized hazard ratio of 0.80. If not, is extension of follow-up or adaptation of the primary outcome from first serious fall injury to all serious fall injuries required? As part of this stage, we will consider the estimate of the cluster design effect needed to determine the inflation to sample size (number of events) required to account for clustering.

Stage 3 concerns issues related to the timing and nature of the interim look.

Proposed timeline

The IMP is intended to assist the DSMB and NIA in making decisions. A proposed timeline is given in the table below subject to the availability of data.

Decision	**Proposed timeline for decision-making**
Change primary outcome definition	Fall 2017 DSMB meeting
Extension of follow-up or adaptation of the primary outcome from first SFI to recurrent SFI	Winter/Spring 2018 DSMB meeting
Interim look	After Fall 2018 DSMB meeting, if extension of follow-up is required and feasible; otherwise, conduct one look at the data at the scheduled end of trial in November 2018.

A flow diagram of the decision-making process is presented in Figure 1.

Stage 1. Changing the Primary Outcome Definition (Fall 2017 DSMB Meeting)

For the primary analysis in STRIDE, a serious fall injury event is defined as a fall accompanied by one or more injuries **requiring medical care.** The injuries are classified as either Type 1 (definitive injuries such as fracture or joint dislocation) or Type 2 (potentially less serious injuries, such as muscle strains). It is expected that Type 1 injuries will require medical attention as a matter of course, while some Type 2 injuries may not.

Since the definition of a fall injury event requires that medical care was sought, any differential referral

to medical treatment in the intervention vs. control arm could result in a difference in apparent incidence of outcome events, even if the overall severity of injuries is the same. Specifically, there is some concern that, through their diligence, **the fall care managers may refer the less definitive of the Type 2 injuries to medical care in the intervention arm than will occur under standard care in the control arm.** We have referred to this as a potential ascertainment bias.

A proposal to change the primary outcome definition to address the potential concern about ascertainment bias was previously circulated to the DSMB (Supplement 1). Specifically, the proposal is to restrict the definition of Type 2 fall injuries (head injury; sprain or strain; bruising or swelling; or other) to require an overnight hospitalization rather than any medical attention. This would leave unchanged all Type 1 injuries (i.e., fracture other than thoracic or lumbar vertebral; joint dislocation; or cut requiring closure) as well as Type II injuries requiring hospitalization. Because of the DSMB's concerns about making a recommendation on this issue after reviewing unblinded data, NIA and PCORI will appoint a working group of the DSMB that will be blinded to intervention assignment. The DSMB will approve its membership. The group will review the outcome event rates under both definitions (current and restrictive) in the control group and, based upon criteria defined in this plan, will make recommendations to NIA and PCORI about changing the primary outcome definition. The DSMB will be informed of the recommendations, but a decision about whether to change the outcome will be made by NIA and PCORI. At the Fall 2017 DSMB meeting, the DSMB, but not its working group advising on the outcome definition, will be provided data by the unblinded statisticians in closed session to assess ascertainment bias.

Assessment of Ascertainment Bias. We will focus on patient-reported type 2 injuries (head injury; sprain or strain; bruising or swelling; or other) without an accompanying type 1 injury. We define Type 2a injuries as those requiring a hospitalization, Type 2b as those that do not require a hospitalization for which other medical attention was sought and Type 2c as those in which no medical attention was sought. Patient-reported events are used for the assessment of ascertainment bias because Type 2c events are not adjudicated. The proportion (p) of Type 2b events in intervention and control groups will be calculated. The denominator for the proportion can be all Type 2 events (2a + 2b + 2c) or only Type 2b and 2c. The difference in proportions with 90% confidence intervals will be calculated and reported. We propose a higher type I error rate for calculation of confidence intervals to provide for greater control of type 2 error because the cost of a missed bias could adversely affect study power even when differences are small. The strategy for assessing the latter is outlined below. Although it is possible that the intervention might "turn a Type 1 event into a Type 2 event," this effect should be negligible relative to the potential ascertainment bias.

Potential Effect of Ascertainment Bias on the Study Power. Using the proportions for intervention (p_i) and control (p_c), we will estimate the "excess" (i.e., due to ascertainment bias) number of Type 2 events in the intervention arm by comparing the observed vs. expected number of Type 2b events. The projected number of excess type 2b events in the intervention arm can be approximated by

> (Observed number of Type 2b events in control) x (hypothesized ratio of events: intervention relative to control) x ($p_i/p_c - 1$).

The derivation is given in Appendix A. The projected excess number of events in the intervention arm will be used to assess its impact on the power of the trial using the observed control event rate and the hypothesized intervention effect. Note, the target number of primary outcomes to detect a hazard ratio of 0.80 with 90% power is 844: 382 intervention vs. 462 control, or a ratio of 0.827. Thus, an excess of 11 events in the intervention arm from 382 to 393, decreases the power to 80% for an annual control event rate of 0.14.

Control Event Rates. If there is evidence of ascertainment bias, we will calculate the observed first event rate in the control arm (while blinded to the overall rate for both groups combined) under both the protocol and modified definitions of the primary outcome. By the fall 2017 DSMB meeting (November-December), the trial will have accrued at least 60% of the total person years of follow-up (PYF, see section 2b and Table A). We assume generation of events will mirror PYF; however, there will be a lag in adjudicated events. Thus, rates will be calculated using patient-reported events and compared with the hypothesized control event rate of 0.14 for 90% power and/or 0.10 for 80% power using the lower 95% exact 1-sided confidence limit.

Proposed Algorithm for Changing Primary Outcome Definition. The proposed algorithm is presented in Figure 1 and briefly summarized as follows:

- No evidence of ascertainment bias: retain original definition
- Evidence of ascertainment bias and control event rate for revised definition > 10%: change to revised definition
- Evidence of ascertainment bias and control event rate for revised definition < 10%: calculate power under diluted intervention effect from ascertainment bias for the protocol definition vs. power under revised definition and choose outcome with least effect on study power.

Stage 2. Evaluation of Event Accumulation and Decision to Extend Follow-up or Adapt the Primary Outcome to Recurrent Events (Winter/Spring 2018 DSMB Meeting)

2a. Monitoring the Cluster Design Effect

Cluster randomized trials require an inflation to sample size (target number of events) to account for the correlation among patients within a cluster. The target number of events in STRIDE was first calculated for an unclustered design and then inflated for interim monitoring and the cluster design effect, which is a function of the intracluster correlation coefficient (ICC) and number enrolled per practice. The number of events for an unclustered design with a hazard ratio (HR) = 0.80 and 90% power is 844. This number was inflated by 1.03 for interim monitoring and 1.52 for the cluster design effect, giving a target number of first primary events of 1321 = 844 x 1.52 x 1.03. The cluster design effect was estimated based on an ICC of 0.0076 from the LIFE trial, assuming clinical practices as clusters, and constant number enrolled/practice of 70 as 1 + (70-1)(0.0076). The design effect needs to be re-estimated for STRIDE because there was wide variability in the number enrolled per practice, and the ICC was estimated from LIFE, which randomized participants and not sites. Because there is no closed form estimate of the ICC for the complexity of the STRIDE trial design (i.e., a clustered survival trial with a competing risk of death), we will estimate the design effect using a variance inflation factor (VIF) method (i.e., variance for the intervention effect for the clustered model relative to the unclustered model), and then use the VIF

to determine the target number of primary events as 844 x 1.03 x VIF. The target number of primary events will then be used to determine information fraction for timing of the interim look. The VIF will be determined by the unblinded trial statisticians as it will involve modeling the intervention effect. The DSMB will be blinded to the actual intervention effect until the scheduled interim look.

2b. Monitoring the Accrual of Adjudicated Primary Outcomes in the Control Group

The primary outcome is first adjudicated SFI. Because adjudication of outcomes will lag behind the accrual of patient reported SFI, an estimate of the conversion rate (with 95% confidence intervals) will be calculated from the accumulated data, i.e., proportion of reported events in which adjudication has been completed (i.e., resolved) that meet the definition of a primary study outcome. Separate conversion rates may need to be estimated for the different types of SFI (Types 1 and 2). The conversion rate will be used to estimate the projected number of adjudicated first SFI at the end of the 40-month trial based on the observed patient reported first SFI rate and the projected total person years of follow-up (PYF) accounting for censoring due to deaths and losses. The observed patient reported first SFI event rate will be estimated as the number of observed patient-reported first SFI divided by the observed total PYF. The projected adjudicated events will then be estimated as:

> Projected adjudicated SFI = observed first patient reported SFI rate x projected total PFY x adjudication conversion rate.

A comparison of the projected number of adjudicated events with the target number of control events (723), after accounting for the estimated ICC, for 90% power (or, if necessary, 540 events for 80% power) will be used to assess whether an extension of follow-up is needed and for how long a duration (see 2c below).

As a supportive strategy, we will also calculate the control event rate. A control event rate of 0.14/year was assumed in the sample size calculations for 90% power; a rate of approximately 0.10/year is needed for 80% power. The sample size calculations also assumed a 7% annual death rate (competing risk) and 3% annual loss to follow-up (censoring) rate. We propose a cumulative incidence rate calculation (with 95% confidence intervals) that considers the censoring and competing risk rate, instead of the Kaplan Meier rate which counts losses and deaths as independent censoring. If the lower bound for the observed rate is greater than the hypothesized control rate of 0.14 for 90% power or 0.10 for 80%, then there is some assurance that the study is on target for 80%/90% power to detect the hypothesized effect, a hazard ratio of 0.80.

2c. Extending Follow-up or Adapting the Primary Outcome from First SFI to All SFI to Maintain Study Power

If the generation of adjudicated control outcomes (first events) is too low to maintain study power at a minimum of 80%, there are two options: 1) extend follow-up for first SFI or 2) adapt the primary outcome from first adjudicated SFI to all adjudicated SFI (i.e., recurrent events). The decision regarding these options needs to be made before an interim look. Extension of follow-up for first SFI is the preferred option because it preserves the primary outcome and has fewer analytical challenges than

adapting to recurrent events. A consideration for extension of follow-up will need to be discussed with the study PIs and funders to determine whether this is feasible.

If extension of follow-up for first SFI is not feasible, adapting the primary outcome will be considered. We will conduct simulation studies to assess power for recurrent SFI using the adjudicated control event rate and the hypothesized intervention effect.

Proposed decision rules if power for first SFI falls below 80% and it is not feasible to extend follow-up assuming a trial of 40 months in duration.

Power for first SFI	**Power for Recurrent SFI**		
	< 50%	**≥ 50%, <80%**	**≥ 80%**
< 50%	Conduct futility analysis (see Stage 3)	Adapt to recurrent SFI*	Adapt to recurrent SFI*
≥ 50%, <80%	Do not adapt	Do not adapt	Adapt if power for recurrent events ≥ 90%*

*If adaptation is done, first SFI will be a secondary study outcome.

Stage 3. Timing and Nature of the Interim Look (At or After Fall 2018 DSMB Meeting)

Based on the above timeline, we propose the following sequence and types of looks at the data.

DSMB Meeting	**Scenario**	**Type of Look**
Winter/Spring 2018 meeting	Power for both first and recurrent events is < 50% under the hypothesized trend and extension of follow-up is not feasible.	Conduct futility analysis based on conditional power for the observed, null and hypothesized trend. If power is < 50% for the observed trend, consider stopping the trial and evaluate trends for secondary outcomes.
Fall 2018 meeting	No extension of follow-up; trial ends as planned after 40 months.	Conduct final analysis
Between Fall 2018 meeting and before final meeting, if necessary	Extension of follow-up is approved	Conduct futility analysis based on conditional power for the observed, null and hypothesized trend. If power is < 50% for the observed trend, consider stopping the trial and evaluate trends for secondary outcomes. Otherwise, continue for one final look at the end of the trial. Note, look follows, and is not part of, the decision to extend follow-up.

Figure 1. STRIDE Decision Algorithm

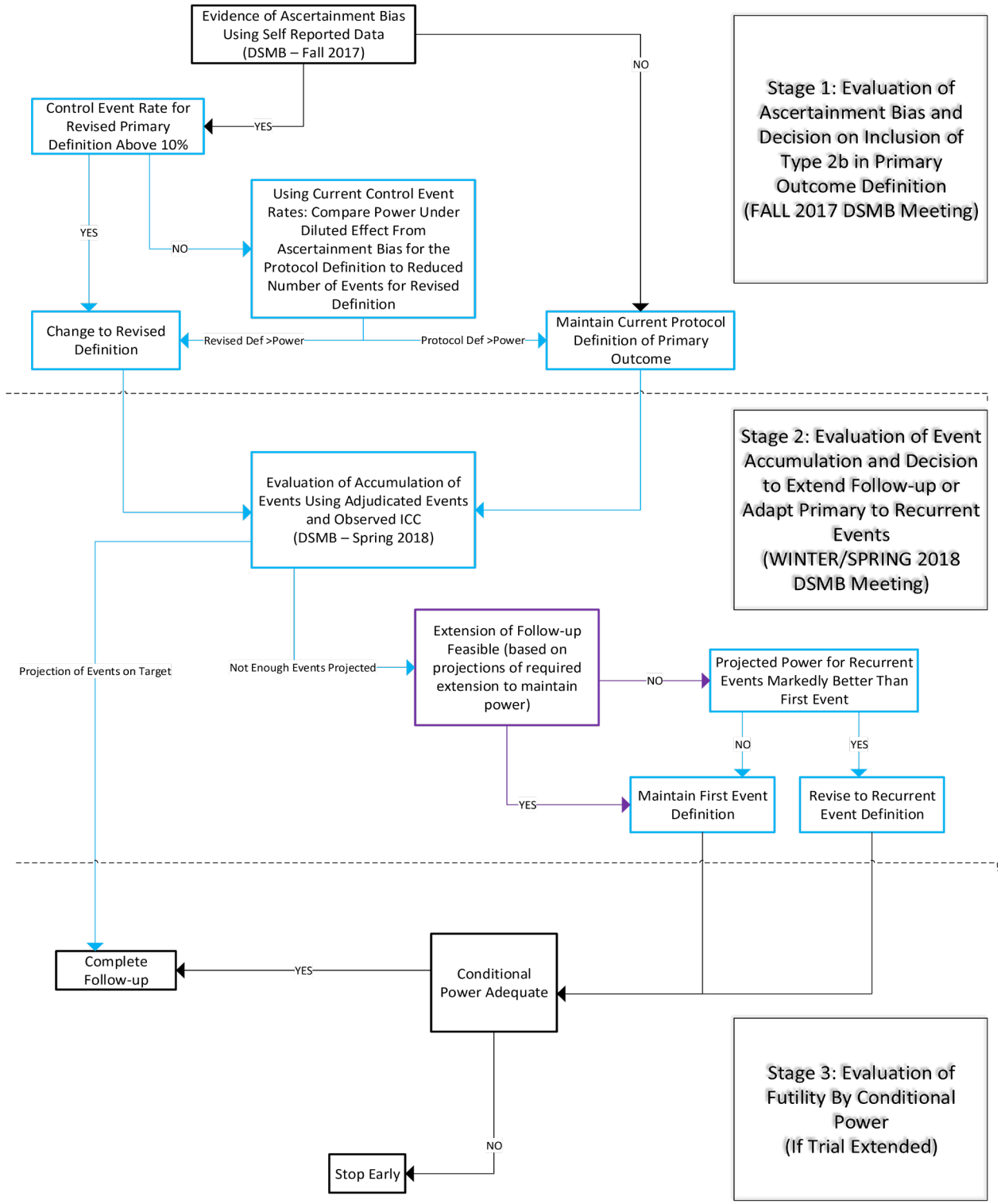


Table A. Projected accrued information fractions for enrollment and RAC capacity variations (based on data through 04/25/17)

Months of total study: 40 months (through Nov 2018)
Months of recruitment: 20 months (through Mar 2017)

end of	accrued PYF	accrued information fraction
Mar17	3846	33.2%
Apr17	4260	36.8%
May17	4683	40.4%
Jun17	5089	43.9%
Jul17	5506	47.5%
Aug17	5918	51.1%
Sep17	6314	54.5%
Oct17	6720	58.0%
Nov17	7109	61.3%
Dec17	7507	64.8%
Jan18	7903	68.2%
Feb18	8257	71.2%
Mar18	8645	74.6%
Apr18	9018	77.8%
May18	9400	81.1%
Jun18	9767	84.3%
Jul18	10142	87.5%
Aug18	10515	90.7%
Sep18	10872	93.8%
Oct18	11238	97.0%
Nov18	11589	100.0%

Based on same assumptions as enrollment/power scenarios

Accrued, not observed, follow-up -- DOES NOT ACCOUNT FOR ASCERTAINMENT AND ADJUDICATIONLAG!

Appendix A

Estimating the "Excess" Number of Type 2 Events in the Intervention Arm Due to Potential Ascertainment Bias

Define Type 2a injuries as those requiring a hospitalization, Type 2b as those that do not require a hospitalization for which other medical attention was sought and Type 2c as those in which no medical attention was sought.

Let P_i = proportion of observed Type 2b injuries in the intervention arm = I_{2b}/D where D could equal $I_{2b} + I_{2c}$ or $I_{2a} + I_{2b} + I_{2c}$.
P_c = proportion in control arm

Let E = excess number of Type 2b injuries in intervention arm.

When there is no ascertainment bias $P_i = P_c$ and then

$$(I_{2b} - E)/D = P_c$$

Note, the denominator is not adjusted by E because we are assuming a shift in events from I_{2c} to I_{2b} because of medical attention; the overall number of events remains unchanged.

Solving for I_{2b}

$$I_{2b} = P_c*D + E \quad (1)$$

From $P_i = I_{2b}/D$ solving for D gives

$$D = I_{2b}/P_i \quad (2)$$

Substituting (2) into (1) and solving for E gives

$$E = I_{2b}*(1 - P_c/P_i) \quad (3)$$

E is the excess number of Type 2b events in the intervention arm potentially due to ascertainment bias under the assumption that the proportion of Type 2b events is the same in intervention and control. The observed number of Type 2b injuries in intervention and the proportions are needed for this calculation.

The excess can also be determined by comparing the observed number of Type 2b injuries in the control arm with the expected number, where the expected number is the number of Type 2b injuries is the observed number in the control arm times the hypothesized intervention effect. The latter would be taken as the hypothesize ratio of events in intervention vs. control (almost equivalent to the hazard ratio). This ratio is based on the target number of primary outcomes in intervention relative to control detect a hazard ratio of 0.80 with 90% power: 382 intervention vs. 462 control, or a ratio of 0.827.

Thus, the excess can be expressed as

$$E = \text{Observed } I_{2b} - \text{expected } I_{2b}$$

Where expected = C_{2b} x hypothesized ratio of events (R = .827). Therefore

$$E = I_{2b} - C_{2b}*R$$

Solving for I2b

$$I_{2b} = E + C_{2b}*R \qquad (4)$$

Substituting (4) into (3) and solving for E gives

$$E = C_{2b}*R*(P_i/P_c - 1) \qquad (5)$$

Thus, either equation (3) or (5) could be used to determine the excess number of Type 2b events in the intervention arm accounting for potential ascertainment bias. Equation (3) uses the observed number of Type 2b events in the intervention arm and equation (5) uses the number in the control arm and the hypothesized ratio of events for intervention relative to control under 90% power = 0.827.

The projected excess number of events can be used to assess its impact on the power of the trial using the observed control event rate and the hypothesized hazard ratio of 0.80.

Supplement 1

Adapting the Primary Outcome Definition in the STRIDE Study
A brief prepared jointly by the Outcomes Committee and the Data Coordinating Center for the STRIDE Study Investigators, and Collated by Joint Principal Investigator Dr. Thomas M. Gill
Submitted: December 5, 2015

Background and Statement of Problem

According to the current STRIDE protocol, the primary outcome of serious fall injury is defined as "a fall leading to medical attention, including non-vertebral fractures, joint dislocation, head injury, lacerations, and other major sequelae (e.g., rhabdomyolysis, internal injuries, hypothermia)".

These fall-related injuries can be classified into two types:

Type 1: fracture other than thoracic or lumbar vertebral; joint dislocation; or cut requiring closure; and

Type 2: head injury; sprain or strain; bruising or swelling; or other.

Over the past 6 months, as the operational details of the STRIDE intervention, particularly those related to interactions between the fall care manager (FCM) and study participants, became more fully developed, a concern was raised about the strong potential for differential bias in the ascertainment of the primary outcome between the intervention and control groups, especially for the Type 2 injuries.

As part of the intervention, participant interactions with the fall care managers (FCMs) could lead to differential medical attention, particularly for the subset of "**type 2"** injuries – head injury, sprain or strain, bruising or swelling, or other – that are insufficiently severe to require a hospitalization. Because receipt of medical care is an inherent part of the primary outcome definition, any factor that differentially causes participants to obtain medical care in one treatment arm versus the other could bias estimates of the difference in outcome rates between the arms. If manifest, this bias would dilute the overall treatment effect, reduce the trial's power, and threaten the validity of the trial's findings.

To address the concern about potential ascertainment bias, the study team has developed a plan to adapt the primary outcome definition.

Proposal to Adapt the Primary Outcome

Under this adaptation, the definition of the Type 2 fall injuries would require an overnight hospitalization rather than any medical attention. Hence, the adapted definition of serious fall injuries would include:

a fall resulting in:

(1) (fracture other than thoracic or lumbar vertebral; joint dislocation; or cut requiring closure) AND any medical attention;

OR

(2) (head injury; sprain or strain; bruising or swelling; or other) requiring hospitalization.

Part 1 of this adapted definition of serious fall injury is unchanged from the original definition.

We considered also including emergency department (ED) visits for the Type 2 injuries (in addition to hospitalization), but decided against it, as described in the **Appendix** under "Alternative Approach".

Rationale for Proposed Adaptation in the Primary Outcome

We are concerned about including fall injuries in which seeking medical attention may be discretionary, since the decision to seek medical attention may be influenced by the participants' interactions with the FCM. For example, as part of the intervention, participants are encouraged to call their FCM about any recent falls, but only after any acute injuries resulting from the fall have been addressed. This notification allows the FCM to modify the treatment plan, if needed. As described in the **Appendix** under, "Plans to Reduce Differential Fall-related Use of Medical Service", we have implemented a plan to provide all participants with a handout on "What to do in case of a fall", which describes the steps that the participant should take in the event of a fall, which may reduce the likelihood of the participants calling their FCM or the Recruitment and Assessment Center (RAC) immediately after a fall. In addition, intervention participants will be advised not to call the FCM for acute injuries arising from a fall. However, it is possible that some patients may still call the FCM in the setting of an injury, and that these interactions may lead participants to seek medical attention with their primary care provider (PCP), at an urgent care center, or an emergency department. In addition, some participants may not know whether some fall-related injuries warrant medical attention and could be prompted to seek medical attention after an interaction with their FCM. Because there would be no such interactions with control group participants, the decision to seek medical attention for some injuries could be differential, biasing ascertainment of the primary outcome, thereby diluting the hypothesized intervention effect size towards the null and reducing the power of the trial.

We consider other serious injuries – i.e. "**type 1"** – fractures, dislocations, or cuts requiring closure or type 2 injuries leading to a hospitalization – to be less susceptible to this potential bias because of their comparatively definitive nature and severity. Because Type 2 injuries can differ greatly in their severity, they are much more susceptible to referral bias. Requiring hospitalization for these Type 2 injuries would greatly reduce any referral bias since the decision to hospitalize the participant should not be influenced by participants' interactions with the FCM. Requiring hospitalization would also ensure that all injuries are serious rather than minor.

However, this adaptation will likely result in a lower first serious fall injury rate (hypothesized to be 0.14 per year in the control group), thereby diminishing the power of the trial. Based on calculations presented under "Statistical Considerations" in the **Appendix**, it appears that, within the range of event rates we consider, power may be more sensitive to ascertainment bias (original definition) than to lower event rates (modified definition); ***thus, preserving the treatment effect is likely a better strategy to optimize power than preserving the event rate*.**

Despite concerns about ascertainment bias, there are no good *a priori* data to estimate the level of this potential bias. The intervention could also shift the distribution of outcome events to less severe injuries (e.g. a fall that might have led to a hip fracture [type 1] results instead in a soft tissue injury [type 2] but not hospitalization) and dilute the treatment effect under the current protocol definition of the primary outcome. Because of these concerns, we present two strategies (primary and supportive) that can be monitored to inform a decision about whether or not the primary outcome definition should be adapted to protect against dilution of the treatment effect. The strategies depend on generating reliable data in a sufficient time frame (prior to the interim look, which is currently scheduled for March 2017) to guide any decision about adaptation.

For this discussion, there are 3 sets of type 2 injuries: (1) those that require hospitalization; (2) those that lead to medical attention but do not require hospitalization; and (3) those that do not lead to any medical attention.

Primary Strategy: Monitor the Event Rate in the Control Arm under the Protocol and Modified Definitions

We would use the accumulating trial data to estimate the event rate in the control arm under both the protocol and modified definitions of the primary outcome. In consultation with the DSMB, we would consider adapting to the modified definition if its corresponding control event rate is comparable to the hypothesized control event rate of 0.14 for 90% power and/or 10% for 80% power (e.g., lower 95% exact 1-sided confidence limit is greater than the hypothesized rate); otherwise, we would retain the current primary outcome definition. Such an adaptation would be based on actual trial data; since only data from the control arm would be evaluated, blinding to treatment effects would be maintained. This strategy would protect against dilution of the treatment effect because of ascertainment bias. It would also provide some assurance that the event rate would be sufficiently high to maintain power if the intervention was effective.

Based on current and projected future recruitment rates, we estimate that there will be approximately 1000 patient years of follow-up in the control group by the end of 2016 or very early in 2017 (see Figure A), after adjusting the time-line to account for a 4-month lag to adjudicate the self-reported fall injuries and those ascertained through utilization data, plus the time required to collect and adjudicate medical records for some events. The exact lower 95% confidence limit of 0.14 for 90% power would be exceeded with an annual control event rate of 0.16, and the limit of 0.10 for 80% power would be exceeded with a control event rate of 0.12 (Table A). These criteria could be used as a basis for determining guidelines for adapting the primary outcome definition. Because the generation of adjudicated outcomes will lag (see **Appendix**, "Time Required to Adjudicate Primary Outcome"), it will be necessary to estimate event rates based both on all patient-reported events and projected adjudicated events. The latter will be calculated using a conversion factor based on the rate of self-reported falls becoming definitive (i.e. adjudicated) serious fall injuries and the rate of definitive serious fall injuries ascertained through utilization data and other non-patient-reported sources of data.

Supportive Strategy: Estimate the Proportion of Nonhospital Type 2 Injuries Leading to Other Medical Attention in each Treatment Arm

The supportive strategy is designed to determine whether ascertainment bias is occurring as postulated. For this strategy, we focus on type 2 injuries (without an accompanying type 1 injury) that do not require hospitalization, and we determine the proportion of these type 2 injuries for which other medical attention was sought. For each fall injury, data are being collected during the every 4-month follow-up interviews about whether the participant saw a doctor or other health care professional for his/her injury and whether the injury resulted in a hospitalization. Based on this information, it will be possible to identify falls that were associated with nonhospital type 2 injuries in the presence or absence of other medical attention. Thus, in principle, all nonhospital type 2 fall injuries, whether or not they resulted in medical attention, can be captured. From these data we could then:

(1) Estimate the proportion of nonhospital type 2 injuries that were associated with other medical attention in the intervention and control arms.
(2) Potentially repeat (1) stratified by the specific type of type 2 injury (e.g. head trauma, soft tissue injury), recognizing that the number of events used in these analyses may be limited.

If the proportion of nonhospital type 2 injuries requiring other medical attention is higher in the intervention group than control group, it would provide some evidence for the postulated bias. The stratification described in (2) would help to guard against an apparent difference because of the intervention (e.g., interaction with the FCM) influencing the distribution of injury types within the type 2 designation.

The supportive strategy involves some degree of accessing randomization information, but it does not involve unblinding, as we will neither compute nor compare event rates in the randomized groups. Instead, it compares the probability of nonhospital medical attention due to type 2 injury, assuming lack of type 1 injury or a prior event, and ignoring person-time.

In summary, the two strategies could be used to inform a decision about adapting the primary outcome definition. The strongest evidence for adapting the primary outcome definition would be if the control event rate under the modified definition was comparable to the hypothesized rate and it was supported by a higher proportion of nonhospital type 2 injuries requiring other medical attention in the intervention arm. It may also be reasonable to adapt based on the former alone without convincing supportive evidence. Adapting based on the supportive evidence alone is less certain, but is potentially worthy of further consideration.

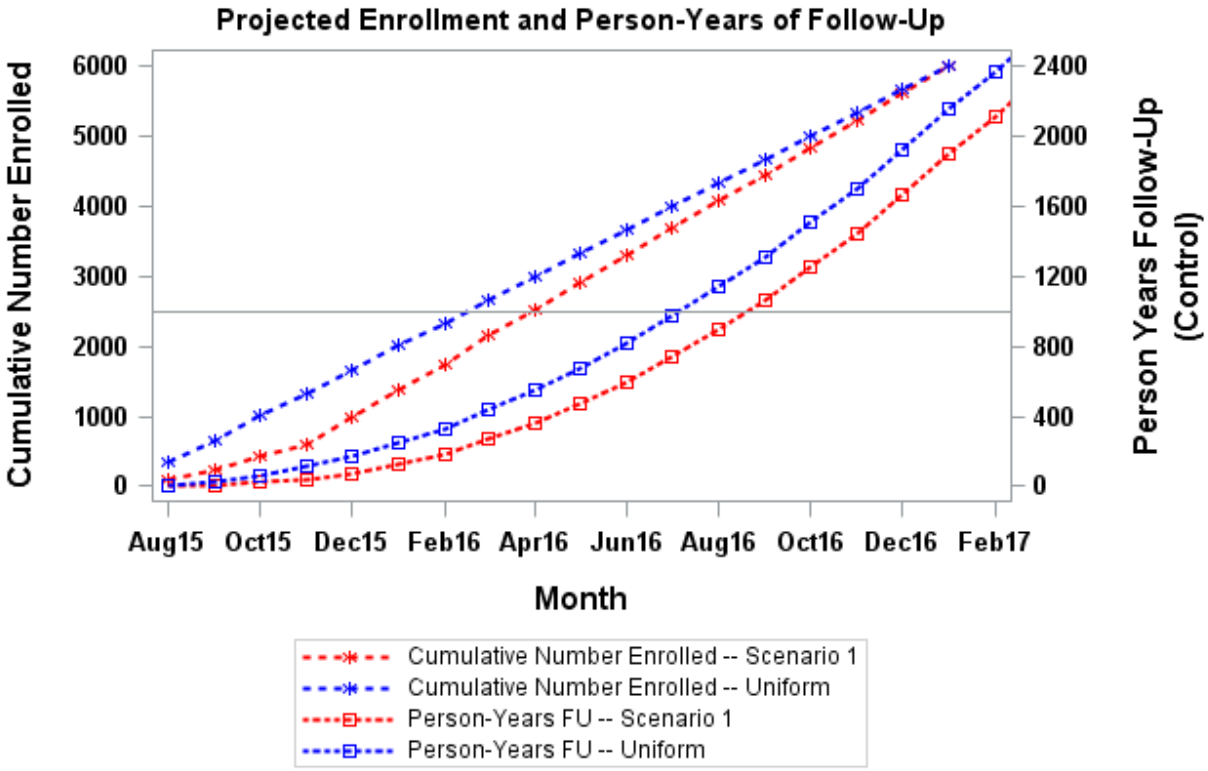
Final Considerations

As described by Schwenk and colleagues,[1] a gold standard for defining a serious fall injury does not exist. Language from the original RFA did not provide an operational definition for serious fall injury. Rather, the RFA indicated that careful attention should be paid to "developing a readily interpretable, operational definition of serious fall-related injuries, such that these events can be ascertained and verified efficiently with a high degree of accuracy."

We believe that our recommendation to adapt the original operational definition of serious fall injury has the potential to reduce referral/ascertainment bias, increase the validity of the trial, and enhance power by preserving the hypothesized treatment effect. Furthermore, Type 2 injuries that result in hospitalization are much more costly and hence important from a policy perspective.

Should the primary outcome be adapted, we also recommend that (1) the original operational definition of serious fall injury be included as a secondary outcome: a fall resulting in (fracture other than thoracic or lumbar vertebral; joint dislocation; or cut requiring closure; head injury; sprain or strain; bruising or swelling; or other) AND any medical attention and (2) Type 1 and Type 2 events be evaluated separately in a secondary analysis.

Figure A



The cumulative number for each month is assumed to be the total number as of the last day of the month. Person years have not been adjusted for the lag related to reconcilation of self-reported falls and those ascertained through electronic means, or the time required to collect and adjudicate medical records.

Uniform recruitment assumes 333 enrolled per month. Scenario 1 uses actual enrollment through the end of November, 2015, and assumes an increase beginning in November, 2015, with equal enrollment/month starting in December, 2015.

Table A. Exact 95% 1-Sided Confidence Intervals for Poisson Rate Based On # Events and Person Years of Follow-up (PY) Observed

rate	PY_1000	PY_1200	PY_1400	PY_1600	PY_1800	PY_2000
0.120	(0.103, --)	(0.104, --)	(0.105, --)	(0.106, --)	(0.107, --)	(0.108, --)
	Event=120	Event=144	Event=168	Event=192	Event=216	Event=240
0.125	(0.107, --)	(0.109, --)	(0.110, --)	(0.111, --)	(0.112, --)	(0.112, --)
	Event=125	Event=150	Event=175	Event=200	Event=225	Event=250
0.130	(0.112, --)	(0.113, --)	(0.115, --)	(0.116, --)	(0.116, --)	(0.117, --)
	Event=130	Event=156	Event=182	Event=208	Event=234	Event=260
0.135	(0.116, --)	(0.118, --)	(0.119, --)	(0.120, --)	(0.121, --)	(0.122, --)
	Event=135	Event=162	Event=189	Event=216	Event=243	Event=270
0.140	(0.121, --)	(0.123, --)	(0.124, --)	(0.125, --)	(0.126, --)	(0.127, --)
	Event=140	Event=168	Event=196	Event=224	Event=252	Event=280
0.145	(0.126, --)	(0.127, --)	(0.129, --)	(0.130, --)	(0.131, --)	(0.131, --)
	Event=145	Event=174	Event=203	Event=232	Event=261	Event=290
0.150	(0.130, --)	(0.132, --)	(0.133, --)	(0.134, --)	(0.135, --)	(0.136, --)
	Event=150	Event=180	Event=210	Event=240	Event=270	Event=300
0.155	(0.135, --)	(0.137, --)	(0.138, --)	(0.139, --)	(0.140, --)	(0.141, --)
	Event=155	Event=186	Event=217	Event=248	Event=279	Event=310
0.160	(0.140, --)	(0.141, --)	(0.143, --)	(0.144, --)	(0.145, --)	(0.146, --)
	Event=160	Event=192	Event=224	Event=256	Event=288	Event=320
0.165	(0.144, --)	(0.146, --)	(0.148, --)	(0.149, --)	(0.150, --)	(0.150, --)

rate	PY_1000	PY_1200	PY_1400	PY_1600	PY_1800	PY_2000
	Event=165	Event=198	Event=231	Event=264	Event=297	Event=330
0.170	(0.149, --)	(0.151, --)	(0.152, --)	(0.153, --)	(0.154, --)	(0.155, --)
	Event=170	Event=204	Event=238	Event=272	Event=306	Event=340
0.175	(0.154, --)	(0.156, --)	(0.157, --)	(0.158, --)	(0.159, --)	(0.160, --)
	Event=175	Event=210	Event=245	Event=280	Event=315	Event=350
0.180	(0.159, --)	(0.160, --)	(0.162, --)	(0.163, --)	(0.164, --)	(0.165, --)
	Event=180	Event=216	Event=252	Event=288	Event=324	Event=360

References

1. Schwenk M, Lauenroth A, Stock C, et al. Definitions and methods of measuring and reporting on injurious falls in randomised controlled fall prevention trials: a systematic review. *BMC Medical Research Methodology.* 2012;12:50.