MrOS Analyst Guide

This analyst guide provides detailed descriptions of variables collected and measures obtained at various MrOS visits. References for variables are also noted. This guide provides information for the following datasets:

V1 (Baseline)
VD (Dental Visit 1)
VI (Interim Questionnaire)
VS (Sleep Visit)
V2 (Visit 2)
V3 (Visit 3)
VI2 (Interim Questionnaire 2)
VS2 (Sleep Visit 2)
V4 (Visit 4)
V4R (Visit 4 Repeat)
VI3 (Interim Questionnaire 3)
V5 (Visit 5)

The visit(s) where the variables were collected are listed.

There are several hundred variables in the MrOS data sets. In many instances there will be a number of comparable variables that could be used in analysis. Often, the decision of which variable to use is driven by the research hypothesis. Here are some basic explanations about the most commonly used variables.

Other Information to Consider
Please note that at the baseline MrOS visit one participant was seen at two different times, about 10 moths apart, and assigned two different IDs (PA3936 and PA3481). The two IDs were followed separately as two different participants until the clinic site realized the problem. Therefore, there were two records for one participant in many of our MrOS datasets. On November 1, 2010 this problem was addressed in the MrOS datasets by deleting PA3936 from datasets and keeping PA3481. The data collected for PA3936 was dropped from all datasets. The only exception is for the baseline QCT scans and data. Baseline QCT scans were only obtained for PA3936. All images and resulting data were kept in the image library and datasets, but the ID was changed from PA3936 to PA3481. All updates made allowed for the most complete and unbiased data. The original source data (e.g., BRI inventory, DNA repositories, CT-scan library, etc.) was updated accordingly whenever possible. The total number of MrOS baseline participants was updated to 5994.

Variables Summarized, in the Order They Appear:

Category	Variables
General Information	Race Variables
Anthropometry	Sitting Height
Anthropometry	Standing Height
Anthropometry	Weight
Blood Pressure	Blood Pressure
Caffeine	Caffeine Use
Cognitive Function	Cognitive Function (3MS, Trails B, Digit Vigilance Test)
Depression	Geriatric Depression Scale
Fatigue	Fatigue Scale
Kyphosis	Kyphosis measurement
Lifestyle	Social Networks
Lifestyle	Life space questionnaire
Lifestyle/QOL	Quality of Life Variables including (SF-12, EQ-5D)
Lung Function	Respiratory symptoms questionnaire
Medical History	Prevalent Fracture
Medical History	Back Pain
Medical History	Hip Pain
Medical History	Rose WHO Questionnaire for angina pectoris
Medical History	Composite Scores for CVD
Medical History	History of atrial fibrillation/flutter and heart rate problems
Medical History	Sexual Function questionnaire
Medications	Medication Use
Moods/Feelings	Goldberg Anxiety and Depression Scales
Physical Activity	Physical Activity Scale for the Elderly (PASE)
Physical Performance	Physical Performance/function (grip strength, walking tests, chair stands, Nottingham Power Rig)
Prostate Health	Prostate health

Category	Variables
Sleep	Sleep General questions
Sleep	Epworth Sleepiness Scale (ESS)
Sleep	Pittsburgh Sleep Quality Index (PSQI)
Sleep	The Functional Outcomes of Sleep Questionnaire (FOSQ)
Sleep	Insomnia Severity Index
Sleep	Restless Leg Syndrome Questionnaire
Specimen Collection	Specimen collection information (fasting information, urine collection)
Tobacco & Alcohol Use	Tobacco/Alcohol use
Vision	Vision Variables
Not released	Frailty Index Information (contact the coordinating center for these variables, and see notes below)

Race variables
Race was collected on the baseline questionnaire.

There are several different ways to examine race in the MrOS data set. At baseline, we asked participants "Which of the following best describes your racial background (Mark all that apply)" Participants could choose from: White; Black or African American; Asian; Hispanic or Latino; American Indian or Alaska Native; Native Hawaiian or Pacific Islander. There are 2 main race variables in the dataset (GIRACE and GIERACE) and one main ethnicity variable (GIHISPA). Since there are several ways to categorize race and ethnicity, in a manuscript that utilizes race, it is imperative that the classification scheme used is described. An additional variable (GINIHRC) that categories race for NIH reporting, is not released but will be made available on a case-by-case basis for NIH reports and grant applications.

GIRACE
This variable has seven race categories: Caucasian/White, African American/Black, Asian, Pacific Islander/Native Hawaiian, American Indian/Alaskan Native, Multiracial and Unknown. 94 Participants listed only an ethnic category (Hispanic) and no race category and are therefore classified as unknown.

GIERACE
This variable has five race and ethnic categories: Caucasian/White, African American/Black, Asian, Hispanic and Other. For this variable, those participants

who marked only "White" are classified as White; those participants who marked only "African American/Black" are classified as African American; those participants who marked only "Asian" are classified as Asian; those participants who indicated Hispanic background are classified as Hispanic (regardless of the race category(ies) selected; and those participants not meeting any of the descriptions above are classified as "Other."

Sitting Height
Sitting height was measured at the baseline visit, dental visit, and visit 2.

Sitting height was measured at the dental visit in Birmingham and Portland. However, the same chair was not used at the baseline dental visit and visit 2. At the dental baseline visit, Birmingham used a stool that measured 46 cm. Portland used two different chairs. If the participant was seen before September 17, 2002 then the chair height was 66.7 cm. If the participant was seen on or after September 17, 2002 then the chair height was 60.3. Chair height was not on the data collection form, but was used for calculating leg length and torso height.

At Visit 2, all clinics used the same stool that measured approximately 45 cm. Chair height was recorded on the data collection form and that recorded height was used to calculate leg length and torso height.

Standing Height
Standing height was measured at the baseline visit, dental visit, sleep visit, visit 2, visit 3, sleep visit 2 and visit 4.

Standing height was measured using Harpenden Stadiometer (Holtain Ltd., Crymych, Dyfed, UK)

The change variables between baseline and each post-visit and between two post-visits were created in the post visit dataset (e.g., change variables between baseline and sleep visit 1 are in the VS1 dataset).

Weight
Weight was measured at baseline, dental, sleep visit 1, visit 2, visit 3, sleep visit 2 and visit 4.

Weight was measured using a balance beam scale, except at the Portland site where weight was measured using a digital scale.

Blood Pressure
Resting Blood Pressure was obtained at Sleep Visit 1, Visit 3, Sleep Visit 2 and Visit 4. Ankle-Arm Blood Pressure was obtained at baseline and visit 3.

For baseline and sleep visit, blood pressure measurements a conventional mercury sphygmomanometer was used to obtain all measurements. Starting with Visit 3, protocols were changed mercury was no longer allowed in most clinical centers. From this time on, the resting blood pressure/pulse measurements and the ankle arm blood pressure measurement were obtained using an automated blood pressure device: the BP Tru automated blood pressure monitor (model BMP-300).

Caffeine Use

Caffeine Use questionnaire data was collected at the sleep visit, visit 2, visit 3 and sleep visit 2.

Please note: this is based on self-report, and is calculated differently than the total caffeine intake based on the Block dietary data from previous visits. The Block data is more complete, based on food frequency questionnaires.

We asked if the participants drank caffeinated coffee, tea, or soda, and if so how many cups per day they used. Based on these answered we calculated caffeine use (mg/day) based on the following formula:

CFCAFF=cups of coffee*136 + cups of tea*48 + cans of cola*36.

If the participant answered no to drinking coffee, a zero was used in the formula. This was also done with the tea and soda variables.

The reference for this is:

Barone JJ, Roberts HR. Caffeine consumption. Food Chem Toxicol. 1996 Jan;34(1):119-29.

SOF comparisons: the SOF study used different caffeine amounts for coffee, tea, and soda. The SOF estimates are based on personal correspondence with Dr Virgina Ernster in 1988. Dr. Ernster gathered many of the same references used by the Barone reference above to create her estimates, but her estimates were never published. We went with estimates for MrOS that were more recent and had a reference we could quote. If you would like to make caffeine intake estimate that is comparable to SOF for cross-study comparisons, use the formula:
Caffeine intake= cups of coffee*95 + cups of tea*55 + cans of cola*45.

At the sleep visit, there are also variables about caffeine use on the night of the in-home polysomnography (POXCOFF, POXTEA, POXSODA, POXCAFF). There are a few participants who have somewhat contradictory data for these variables and the “CF” variables. There are 18 participants who say they do not currently drink caffeinated coffee (CFCCOF) but report dinking it the night of the polysomnography, 37 participants who report not drinking caffeinated tea

(CFCTEA) but report drinking it the night of the polysomnography(POXTEA), and 44 participants who report that the did not drink caffeinated sodas(CFCCOK), but report drinking it the night of the polysomnography (POXSODA). A variable summarizing Caffeine use is available as of the FEB19 release, called POXCAFF.

Cognitive Function:

Teng Modified Mini-Mental State Exam (3MS)
The 3MS was collected at the baseline visit, sleep visit, visit 2, visit 3, sleep visit 2 and visit 4.

The 3MS is a test to assess cognitive abilities. The range of the score is from 0 to 100, with higher scores representing better cognitive function (TMMSCORE). There are some subscores also (TMNAMING, TMRECALL, TMREGIS, TMREVERS, TMSPACE, TMTEMPOR). The change variables between baseline and each post-visit, and between two post-visits were created in the post visit dataset (e.g., change variables between baseline and visit 2, and sleep visit 1 and visit 2, were all included in the V2 dataset).

The reference for this is:

Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. J Clin Psychiatry. 1987 Aug;48(8):314-8.

If similar variables are needed as those collected in the SOF study contact the San Francisco Coordinating Center. Comparable variables to the SOF variables SHT3MS (the 26-point modified mini-mental state exam) or MMSE (the 30 point MMSE) can be calculated.

Trail Making B Test
The Trail Making B Test was collected at the baseline visit, sleep visit, visit 2, visit 3,sleep visit 2 and visit 4.

The Trail Making B Test is a timed test that measures attention, sequencing, visual scanning and executive function. A faster time for completion (in seconds) represents better cognitive functioning.

Since August 2009 data release, protocol violations and possible errors were discovered in the data from baseline visit, sleep visit, and visit 2. Sites reviewed forms and when possible, data was updated. When corrections could not be determined the data was set to .W=weird value or .A=missing.

The reference for this is:

Reitan R. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 1958;8:271-276.

The analysis variable is TBSECON, which ranges from 0-300.

If the similar variable is needed as that collected in the SOF study contact the San Francisco Coordinating Center. The Comparable variable to the SOF variable TRLBTS (the extrapolated trails B test time) can be calculated.

The protocols for all post-baseline visits state that participants were only allowed 5 errors (TBERROR). If they had 5 errors the test was to be stopped and the completion time (TBSECON) was to be set to 300 seconds. At baseline there was no restriction on the number of errors. The variable TBSECON was set to 300 if TBERROR was 5 or more at the baseline visit, to make it comparable to the protocols at the following visits.

The TBSECON, total time to complete the trails B test (seconds), variable had 2 problems at the baseline visit.

The first problem was that some clinic staff at the baseline visit were recording the time in minutes-seconds rather than seconds (for example, 1 minute 10 seconds should be recorded as 70 seconds but it was being recorded as 110). A memo was sent out to the clinics in October of 2000 and staff were asked to correct any known errors in the data system. All data for clinic visits before November 2000 were examined within each staff ID to look for data patterns that would indicate that the data were recorded as minutes-seconds rather than seconds (the data collected incorrectly as minutes seconds does not have any times between 60 and 99, 160 and 199, and 260 and 299 seconds). All data that were suspected to be incorrectly collected were queried and corrections were made whenever possible. We also looked at frequencies of this total time variable within each staff ID by month to see if the problem continued after 10/2000.

The second problem at baseline was that the test was to be administered for 300 seconds, but it appeared that some staff IDs might have stopped the test early at 180 seconds (3 minutes). This confusion may be because other protocols, such as the SOF study, use 180 seconds as the cut-off time. All data was examined within staff ID looking for “lumps” at time points for early cut-off or rounding error. We also looked at the frequencies of TBSECON within staff ID by month to see if there was a pattern of “lumpiness” that was corrected as time went on. Five staff IDs appeared to have “lumpy” data. We queried the sites regarding these participants with potential early cut-off times. It was confirmed that 29 participants had a known early cut-off of 180 seconds, so TBSECON was set to .W (weird value). For the others with potential early cut-off times that we could not resolve, we decided to set participants who had TBSECON=180 and the number of circles competed<25 (meaning they did not finish the test) to .W for those 5 staff IDs in question. This effected 26 participants.

The third problem at baseline was that that the Palo Alto clinic did not follow the protocol for recording number of errors during the baseline visit. The protocol stated that each time a participant made an error (that is, drew a line to the wrong circle) the interviewer should: 1) stop the participant immediately after the error is made, 2) draw a mark through the mistaken line, 3) record the error, and 4) direct the participant back to the last circle reached correctly. The Palo Alto clinic only reported the number of "protocol errors" encountered; for example, they recorded an 'error' when the participant wasn't immediately told of a mistake in the connection of the circles or the interview was interrupted, etc. Therefore, the Palo Alto clinic has a very low error rate; they have 955 participant with "0" errors, while the average number participants without any errors for the other clinics is about 575 participants. We suggest that the number of errors should not be used in analysis; only the total time to complete the test should be analyzed.

Compared to publications using men of comparable age, our means, standard deviations, and ranges are similar (data for Caucasian men only). [1,2]

Yaffe K, Lui L, Zmuda J, Cauley J. *Sex Hormones and Cognitive Function in Older Men*. JAGS 50:707-712, 2002.

Barrett-Connor E, Goodman-Gruen D, Patay B. *Endogenous Sex Hormones and Cognitive Function in Older Men*. J Clin Endicrinol Metab 84: 3681-3685, 1999.

Digit Vigilance

The Digit Vigilance test was performed at the sleep visit and sleep visit 2.

The digit vigilance test is a test of sustained attention and psychomotor speed. It is a paper-and-pencil task designed to measure vigilance during rapid visual tracking and accurate selection of target stimuli. The task appears to isolate alertness and vigilance and to place minimal demands on the selectivity and capacity components of attention. The test is part of the Repeatable Cognitive Perceptual Motor (RCPM) Battery. Participants are asked to cross out as quickly as possible each ‘6’ that is followed by a number greater than 6 (7,8,9) that appear randomly within 59 rows of 35 digits. **NOTE:** the standard test instructions say the participant is to cross out all 6’s, regardless of what number they are followed by. The distinction should be made that the MrOS version of the test is more difficult than the standard DVT.

Total time (DVTMSCOR) to complete the test is gathered, with higher scores indicating greater levels of impairment. The number of omission errors (6’s not marked) is gathered (DVOMERR) as well as the number of commission errors (DVCOMERR=number marked other than 6, 6’s marked that are not followed by 7, 8 or 9). These scores were summed to create total errors (DVTOTERR). For

DVOMERR, DVCOMERR and DVTOTERR a lower score represents better attention and psychomotor speed.

Empirically, total time is a better substantiated measure than total errors and is the primary measure derived.

The references for the test are:

Lewis RF, Rennick PM.(1979) Manual for the Repeatable Cognitive-Perceptual-Motor Battery. Gross Point Park, MI: Axon Publishing.

Lewis RF. (1995). Digit Vigilance Test Professional User's Guide. Psychological Assessment Resources, Inc. Lutz, FL.

Geriatric Depression Scale (GDS)

The Geriatric Depression Scale was asked at the sleep visit, visit 2, visit 3, interim questionnaire 2, sleep visit 2 and visit 4.

This is a standard scale used to measure depression, ranging from 0 to 15 (DPGDS15). The variable DPGDS15=Geriatric Depression Score 15 point scale was calculated using the 15 yes/no questions found on the form for "Moods in Past Week".
There is also a standard cutpoint used to define depression yes/no of GDS>=6 (DPGDSYN).

The standard references for this are:

Sheikh J, Yesavage J. Geriatric Depression Scale: recent evidence and development of a shorter version. Clin Gerontol.1986; 5:165-173.

Almeida OP, Almeida SA. Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. Int J Geriatr Psychiatry 1999; 14:858-865.

Fatigue Scale

A fatigue scale was collected as part of sleep visit 2.

A set of 6 questions related to fatigue were asked. The same set of questions were previously asked at Year 5 of Health ABC and the SEA Pilot study. Variables include: SLTIRE, SLWKLEV, SLSLPLEV, SLLIVLEV, SLTIRLEV and SLENRLEV. The range for the last five variables is 0-10 where is 0 is 'not at all' and 10 is 'very' (i.e. for SLWKLEV which asks 'During the past month, how weak did you fell' 0 is not weak at all and 10 is very weak.

Pittsburgh Fatigability Scale

The Pittsburgh Fatigability Scale was administered at Visit 4.

Two main summary score variables were created, PFPHYSSC for physical fatigability, and PFMENTSC for mental fatigability. A higher score means greater fatigability.
A reference for the Pittsburgh Fatigability Scale is:

> Glynn NW, Santanasto AJ, Simonsick EM, Boudreau RM, Beach SR, Schulz R, Newman AB. The Pittsburgh Fatigability scale for older adults: development and validation. J Am Geriatr Soc. 2015 Jan;63(1):130-5. doi: 10.1111/jgs.13191.

Kyphosis Measurement
A kyphosis measurement was obtained as part of visit 3.

To measure kyphosis, small wooden blocks (1.7 cm thick) were place under the participants head until the participant's head is in a neutral position. The number of blocks required is variable KYBLOCKN. The range is 0-10. If more than 10 blocks were required, 10 was recorded on the data collection form.

Please note that all staffs were certified to perform this measure. However, some data was collected prior to certification. The flag variable KYPOST indicates if data was collected pre or post certification.

Please note that the kyphosis measurement was added a couple of months after Visit 3 started. Each clinic site added the kyphosis measurement at different time point. The kyphosis measurements were assigned A:Missing for all participants who had their Visit 3 date before the kyphosis measurement was added to the visit. Participants with visit type as "SAQ only" or "Home visit" were also assigned as A:Missing for the kyphosis measurements.

The reference for this measure is:

> Kado, DM, Huang MH, Karlamangla A, Barrett-Connor E, Greendale GA. Hyperkyphotic posture predicts mortality in older community dwelling men and women: a prospective study. J Am Geriatr Soc 2004; 52:1662-1667.

Social Networks
Questions related to social networks were collected at the interim questionnaire, visit 2, visit 3 and visit 4.

The following reference provides more information.

> Michael YL, Berkman LF, Colditz GA, Kawachi I. Living Arrangements, Social Integration, and Change in Functional Health Status. American Journal of Epidemiology. 200. Vol. 153, No. 2: 123-131.

Life space questionnaire
Life space questions were asked at visit 3, interim questionnaire 2 and visit 4.

Participants were asked about their mobility on 5 different life space levels (within home, outside house, within neighborhood, within town, outside town), taking into account the frequency of movement and degree of independence during the movement. The seven calculated variables from this questionnaire are: maximal life space (LSM), assisted life space (LSE), independent life space (LSI), restricted life space (LSID), measure of level & independence (LSII), measure of level & frequency (LSIII), and composite life space score (LSC).

The method used to recode inconsistent data and handle missing values is consistent with the suggestion from the reference listed below:

- If a participant indicated they went to a higher level, data was corrected so that he also went to the lower level
- if the frequency of how often a participant went to a lower level is less than that of a higher level or if it is missing, data for the lower level was set to that of higher level
- if the independence indicator is missing for a higher level, it was set to that the same as the lower level
- if “personal assistance” was indicated for a lower level, it was also indicated for the higher level
- if “equipment only” was indicated for a lower level, and if “no assistance” was indicated for a higher level, “equipment only” was set for the higher level.

The references for this are:

Claire Peel, Patricia Sawyer Baker, et al. Assessing mobility in older adults: The UAB study of aging life-space assessment.
Physical Therapy. Vol. 85, No. 10: 1008-1019, 2005.

Patricia S. Baker, Eric V. Bodner, Richard M. Allman. Measuring life-space mobility in community-dwelling older adults.
JAGS. Vol. 51, No. 11: 1610-1614, 2003.

Quality of Life variables, including SF-12 and EQ-5D

Quality of Life variables were collected at baseline, interim questionnaire, sleep visit, visit 2, visit 3, sleep visit 2 and visit 4.

The EQ-5D was asked at the interim questionnaire, visit 3,sleep visit 2, and visit 5.

The main variables to use from the Quality of Life Questions from the Self-Administered Questionnaire are QLCOMP, QLFXST52, QLMCS12 and QLPCS12. The QLMCS12 and QLPCS12 are summary score variables from the SF-12 questionnaire. The change variables between baseline and each post-visit, and between two post-visits were created in the post visit dataset (e.g.,

change variables between baseline and sleep visit 1, and interim 1 and sleep visit 1, were all included in the VS1 dataset).

Please note that the variables QLMCS12 and QLPCS12 were not able to be calculated at the interim questionnaire because the 3 variables QLBLUE, QLCALM, QLENERGY found at baseline and sleep are collected slightly differently here. On this interim questionnaire there were only 4 categories for these variables, rather than the 5 categories required by the formula to create the calculated variables QLMCS12 and QLPSC12. Therefore, these variables have different names (QLBLUEI, QLCALMI, QLENERGI).

Please note that the SF-12 we use in MrOS is modified from the validated version. The question use in MrOS is: "Compared to other people your own age, how would you rate your overall health?" has the options Excellent for my age, Good for my age, Fair for my age, Poor for my age, Very Poor for my age. The question in the validated SF-12 is "In general, would you say your health is: Excellent, Very Good, Good, Fair, Poor."

The suggested citation for the SF-12 is:

> Ware JE, Kosinski M, Keller SD. SF-12: How to score the SF-12 Physical and Mental Health Summary Scores. Lincoln, RI: QualityMetric Incorporated, Third Edition, 1998.

From the reference:

> "The SF-12 is a multipurpose short-form generic measure of health status. It was developed to be a much shorter, yet valid, alternative to the SF-36 for use in large surveys of general and specific populations as well as longitudinal studies of health outcomes."

Another reference for the SF-12 is:

> Ware JE, Kosinski M, Keller SD. A 12-Item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220–33.

Both the QLPCS12 and the QLMCS12 are scored using norm based methods. Physical and mental regression weights and a constant for both measures come from the general US population. Both the QLPCS12 and QLMCS12 are transformed to have a mean of 50 and a standard deviation of 10 in the general US population. HOWEVER, please note that since MrOS uses a modified version of the SF-12, the comparisons of SF-12 data across studies might not be valid.

QLPCS12

Derived from the SF-12, this is the Physical Health Summary Measure. It is composed of 4 subcategories: Physical Functioning (QLMODLIM and QLSEVLIM); Role-Physical (QLACCOM and QLKIND); Bodily Pain (QLPAIN); and General Health (QLHEALTH).

QLMCS12
Again from the SF-12, this is the Mental Health Summary Measure. It is comprised of 4 subcategories: Vitality (QLENERGY); Social Functioning (QLSOCIAL); Role-Emotional (QLACCLV and QLCARE); and Mental Health (QLCALM and QLBLUE).

QLCOMP
A calculated variable, QLCOMP classifies those with excellent/good health (1) vs. those with fair/poor/very poor health, from QLHEALTH. This calculated variable is not a component of the SF-12 (although QLHEALTH is part of the modified SF-12.)

QLFXST52
A calculated variable that indicates the total difficulty with all five IADLs.

Additional references:

Fitti JE, Kovar MG. The supplement on aging to the 1984 National Health Interview Survey. Vital & Health Statistics-series 1: Programs & collection procedures. 1987;21:1-115.

Pincus T, Summey JA, Soraci SA Jr et al. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983;26(11):1346-1353.

The EQ-5D is an internationally recognized measure for quality of life estimation. There are 5 questions that ask about mobility, self-care, usual activites, pain/discomfort and anxiety/depression.

Visit 5 participants were also asked to rate their health on a scale of 0 -100, where 0 is the worst state you can imagine and 100 is the best state you can imagine. This is variable QLSCALE

Note: The EQ-5D may not be reproduced in a publication or on the internet without prior written permission from the EuroQol Office (https://euroqol.org).

The following references can be used:

- EuroQol Research Foundation. EQ-5D-3L User Guide, 2018. Available from: https://euroqol.org/publications/user-guides.
- Johnson JA, et al. Valuation of EuroQOL (EQ-5D) health states in an adult US sample. Pharmacoeconomics 1998;13:421–33.

Respiratory Symptoms Questionnaire
The respiratory symptoms questionnaire was collected at sleep visit 2.

Several variables from the ATS-DLD-78 Adult Questionnaire (American Thoracic Society for the Division of Lung Diseases) were asked. Please note not all the questions from this questionnaire were included in MrOS. These variables start with the two letter code ‘LF’ for Lung Function.

The following references are for the respiratory symptoms questionnaire.

Ferris BG. Epidemiology Standardization Project (American Thoracic Society). Am Rev Respir Dis. 1978 Dec;118(6 Pt 2):1-120.

Helsing KJ, Comstock GW, Speizer FE, Ferris BG, Lebowitz MD, Tockman MS, Burrows B. Comparison of three standardized questionnaires on respiratory symptoms. Am Rev Respir Dis. 1979 Dec;120(6):1221-31.

Prevalent Fracture
There are several variables which can be derived from the detailed Fracture History Information collected at the baseline visit.

There are fourteen calculated variables for prevalent fracture definitions. These are:

FFNOHP	Prevalent non-hip fracture (at any age)
FFNOHS	Prevalent non-hip, non-spine fracture (at any age)
FFNOHSW	Prevalent non-hip, non-spine, non-wrist fracture (at any age)
FFNOSP	Prevalent non-spine fracture (at any age)
FFNT502	Any non-traumatic fracture after age 50 (yes/no)
FFNT504	History of non-traumatic fracture (categorized as never (0), after age 50 (1), at/pre age 50 (2), traumatic fracture at any age (3))
FFNTGT50	Number of non-traumatic fractures after age 50
FFNTLE50	Number of non-traumatic fractures at or before age 50

The "after age 50" variables are calculated for comparison with SOF. In SOF, fractures after age 50 were generally considered to be postmenopausal. Use of these variables for the men in MrOS will depend research question.

Back Pain
Questions regarding back pain were asked at baseline, visit 2, visit 3, and visit 4

Baseline included general back pain questions as well as an extended set of questions that were adapted from the North American Spine Society (NASS) questionnaires for back and neck pain. Visit 2 only included general back pain questions.

References for the extended set of back pain questions include:

Daltroy LH, Cats-Baril WL, Katz JN, Liang MH. The North American Spine Society lumbar spine outcome assessment instrument, reliability and validity results. Spine 1996;21:741-748

Compendium of Outcome Instruments for Assessment & Research of Spinal Disorders, ed Gatchel RJ. North American Spine Society, LaGrange IL, 2001.

Hip Pain
Hip pain questions were asked at visit 2 and visit 4.

The four calculated variables from this questionnaire are: standard WOMAC pain score on the right hip (BHWPSR), standard WOMAC pain score on the left hip (BHWPSL), modified WOMAC pain score on the right hip (BHWPMR), and modified WOMAC pain score on the right hip (BHWPML).

In MrOS, WOMAC pain questions are asked about the hips(eight questions. The modified hip pain subscale was calculated based on these 8 questions. The WOMAC hip pain (modified) subscale calculation was modified from the code of MOST study which was created by Yun Yi Hung (Coordinating Center).

There is also a “don’t do” response option included for “Going up or down stairs” question. If the participant chose the “don’t do” response, the score for that question was set to missing when computing WOMAC scores.

The method used to handle missing values (ie., participant fails to/refuses to complete all questions) is consistent with the suggestion from the WOMAC User’s Guide (Nicholas Bellamy) for how missings should be treated: “If >= two pain, both stiffness, or >= four physical function items are omitted, the patient’s response is regarded as invalid and the deficient subscale(s) should not be used in analysis. Where one pain, one stiffness, or 1-3 physical function items are missing, we suggest substituting the average value for the subscale in lieu of the missing item value(s).

The reference for this scale is:
Nicholas Bellamy, WOMAC Osteoarthritis Index, User guide V.

Information about the WOMAC Osteoarthritis Index can be found at the following website: http://www.womac.org

Rose WHO Questionnaire for Angina Pectoris
The Rose Questionnaire was asked at the sleep visit and sleep visit 2.

This questionnaire is commonly used to determine the prevalence of angina in epidemiologic studies. This definition of angina (CVROSE=1) is based on

established criteria: chest pain that comes on with exertion, that causes a person to stop or slow down and goes away within 10 minutes, and is located over the sternum or in both the left chest and left arm. There is also a severity grade (CVROSEGR) which can be 1 or 2. (1= getting the chest pain when they walk uphill or in a hurry, 2=getting the pain when uphill or in a hurry, but also when walking at an ordinary pace on level ground).

The standard reference for this is:

> Rose G, Blackburn H, Gillum RF, Prineas RJ. Cardiovascular Survey Methods. 2nd Edition. Geneva: WHO; 1982

There is also a Rose score for intermittent claudication. We were unable to calculate this because our forms are missing 2 of the 8 questions needed to calculate the score.

Composite Scores for Cardiovascular Disease

The composite scores for cardiovascular disease were calculated at the sleep visit and sleep visit 2.

The composite scores include CVCHD (Had previous diagnosis of coronary heart disease?), CVCER (Had previous diagnosis of cerebrovascular disease?), and CVPVD (Had previous diagnosis of peripheral vascular disease?). The event group for each of the composite score is:

CVCHD:

- MHMI (Heart attack)
- CVCABG (Heart bypass)
- CVAPCORA (angioplasty of coronary arteries)
- MHANGIN (angina (chest pain))

CVCER:

- MHSTRK (stroke)
- CVTIA (Transient ischemic attack)

CVPVD:

- CVBLKA (Intermittent claudication)
- CVAORANE (Repair of aortic aneurysm)
- CVBPLEGS (Bypass procedure on the arteries of legs)
- CVAPLOW (angioplasty of lower extremity arteries)
- CVSURGBV (Carotid endarterectomy)

History of Atrial Fibrillation or Flutter, Heart Rate Problems

Gathered at the sleep visit and sleep visit 2.

The primary source of the variables MHAFIB and MHHR are from the ECG worksheet. If the ECG worksheet data could not be found then similar data from the PSG Study Evaluation form was gathered. The variable MHAFIBS notes the source of the MHAFIB data and MHHRS notes the source of the MHHR data (ECG worksheet or PSG Study Evaluation form). Most data comes from the ECG

worksheet (87%). Of note, a much larger percentage of data came from the PSG Study Evaluation form for the Pittsburgh (PI) site compared to the other clinics. Those with missing data for the variable will have a value of .M=not applicable for the source variables (MHAFIBS and MHHRS). The values for MHAFIB and MHHR are .A=Missing if the forms are both missing or if the form data was in the system but that variable has missing data. The value for MHAFIB and MHHR is .M=not applicable if the participant did not have an ECG measurement. Note: a few men do have data for this variable that did not have an ECG done (Sleep visit 1 for PI5492, PO6857, PO7420).

Differences in Wording on the 2 forms:

Variable	Wording on ECG Worksheet	Wording on PSG Study Evaluation
MHAFIB	Have you ever been diagnosed with atrial fibrillation or atrial flutter?	At the clinic visit, did the participant report known atrial fib flutter?
MHHR	Has a doctor ever told you that you have problems with your heart rate?	At the clinic visit, did the participant report other known HR problem?

Sexual Function Questionnaire
Sexual Function questions were asked at visit 2 and visit 3.

Two questions on erectile dysfunction have been asked at visit 2, visit 3 and interim questionnaire 2 (SFEDYSF, SFTRBERE).

Participants were asked to complete a short sexual function questionnaire. Participants were not required to answer any question. They could leave any or all blank. Clinic staff did not review the questionnaires for completeness. The questionnaires were submitted to the data system as is and no edits were posted to the data. No additional data checks were performed.
Flag variable SFFLAG has been created to indicate if a participant completed any part of the Sexual Function Questionnaire.

SHIM SCORE and Erectile Dysfunction Calculated Variables

SFSHIM
SFEDSHIM

An abridged 5-item version of the International Index of Erectile Function (IIEF-5) Questionnaire was administered as part of the sexual function questionnaire. This is commonly referred to as the Sexual Health Inventory for Men (SHIM). The main reference does not have a category 0-No Sexual Activity/Did not attempt intercourse and range is 5-25. However, other papers that use SHIM

have the same categories that were used in MrOS and also have the 0=No sexual activity/did not attempt intercourse category and range is 1-25. Higher values represent better functioning. SFSHIM has a range of 1-25. Literature show a cutpoint of <= 21 as having erectile dysfunction. SFEDSHIM indicates if the participant has erectile dysfunction. No reference mentions what to do if the participant is missing one or more of the answers, so we set those to missing if all 5 questions are not answered.

The standard reference for this is:

> Rosen RC, Cappelleri JC, Smith MD, et al. Development and evaluation of an abridged, 5-item version of the International Index of Erectile Disfunction (IIEF-5) as a diagnostic tool for erectile dysfunction. Int J Impotence Res 11:319-326, 1999.

Medication Use

Medication variables were collected at the baseline visit, interim questionnaire, sleep visit, visit 2, visit 3, interim questionnaire 2, sleep visit 2 and visit 4.

Medication variables from the baseline visit are included in the V1 dataset and the M1 dataset. Originally, medications were not collected using the medication inventory forms. These medications variables all start with the 2 letter prefix code MU and can be found in the V1 dataset. Medication information that was collected at the baseline visit, was later entered using the medication inventory forms used at follow-up MrOS visits. These data are in a separate baseline medication dataset (M1). At all visits after baseline, medication data were collected using the medication inventory forms and data can only be found in the medication data sets (MI, MS, M2,M3, MS2, M4). Extra medication variables can be found in the MF datasets (MF1, MFI, MF2, MF3, MFI2, MFS, MFS2 and MF4). Please see the medication data information files for more.

Please note, the M1 dataset is the preferred dataset that should be used for analyses.

A reference used for using coded medication data is:
Pahor M, Chrischilles EA, Guralnik JM et al. Drug data coding and analysis in epidemiologic studies. Eur J Epidemiol. 1994 Aug;10(4):405-11.

A description for the methods section is:
All prescription medications recorded by the clinics were stored in an electronic medications inventory database (San Francisco Coordinating Center, San Francisco, CA). Each medication was matched to its ingredient(s) based on the Iowa Drug Information Service (IDIS) Drug Vocabulary (College of Pharmacy, University of Iowa, Iowa City, IA).

Please see the medication database analyst guides for more information. In particular, the document MU_M1_compare.pdf compares the medication data at baseline.

Goldberg Anxiety and Depression Scales

The Goldberg Anxiety and Depression Scales were asked at sleep visit, visit 2 and sleep visit 2.

This shortened scale was derived by latent trait analysis from a standardized psychiatric research interview. It is intended for use by non-psychiatrists in clinical investigations. The anxiety scale (AXANXSC) ranges from 0 to 9, with higher scores representing more anxiety. A standard cutpoint for clinically important disturbance for anxiety (>=5) is also created (AXANX50). The depression scale (AXDEPSC) ranges from 0 to 9, with higher scores representing more depression. A standard cutpoint for clinically important disturbance depression (>=2) is also created (AXDEP50).

The main reference for this is:

Goldberg D, Bridges K, Duncan-Jones P, Grayson D 1988 Detecting anxiety and depression in general medical settings. Bmj 297(6653):897-9.

Personality Questionnaire

The personality questionnaire is a 5 page questionnaire. Each question get answered with a 1 through 5 rating. 7 summary score variables are created by summing up selected questions, some of which will be in reverse order.

PEOPSCOR: Optimism summary score (5-30)
PECOSCOR: Conscientiousness summary score (5-50)
PETASCOR: Trait Activity summary score (8-40)
PEGDSCOR: Goal disengagement summary score (4-20)
PEGRSCOR: Goal re-engagement summary score (6-30)
PEGDAVG: Goal disengagement average score (1-5)
PEGRAVG: Goal re-engagement average score (1-5)

The main references are:

Wrosch C, Scheier MF, Miller GE, Schulz R, Carver CS. Adaptive self-regulation of unattainable goals: goal disengagement, goal reengagement, and subjective well-being. Pers Soc Psychol Bull. 2003;29(12):1494-508.

Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. J Pers Soc Psychol. 1994;67(6):1063-78.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), Personality Psychology in Europe, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences (http://ipip.ori.org/). Internet Web Site.

Physical Activity Scale for the Elderly (PASE) variables
The PASE was collected at baseline, sleep visit, visit 2, visit 3, interim questionnaire 2,sleep visit 2 and visit 4.

The Physical Activity Scale for the Elderly has been developed by the New England Research Institute. PLEASE NOTE THAT THE PASE QUESTIONNAIRE IS COPYRIGHTED MATERIAL AND CANNOT BE REPRODUCED WITHOUT THE EXPRESS CONSENT OF NERI.

PASE scores are summary values calculated from weights and frequencies for each of the 12 types of activities described in the questionnaire. For more information about the PASE scores, please see the cleaning code from the baseline exam and the following reference:

> Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): Development and Evaluation. Journal of Clinical Epidemiology. Volume 46, Number 2. Pages 153 -162.1993.

PASCORE
This is the final summary score for the PASE questionnaire. Note that the sitting activities do not contribute to the summary score. There is not a range for this variable. The range for the given visit (or analysis dataset) can be reported in manuscripts.

PASE subgroups

The subscores for each of the 12 types of activities are as follows:

PACAREW	Caring for another person
PAGARDNW	Outdoor Gardening
PAHHWW	Heavy housework
PAHOMEW	Home Repairs
PALAWNW	Lawn work/yard care
PALHWW	Light housework
PALTEW	Light sport/rec activities
PAMODW	Moderate sport/recreation activities
PASTRW	Strenuous sport/recreation activities

PAWALKW	Walking activities
PAWGTW	Muscle Strength/Endurance activities
PAWKW	Work for pay/volunteer

PASE subscores for leisure exercise (PASELEIS), household activity (PASEHOUS) and occupational activity (PASEOCC) were also created based on the groupings described in this reference:

> Washburn RA, Ficker JL. Physical Activity Scale for the Elderly (PASE): the relationship with activity measured by a portable accelerometer. J Sports Med Phys Fitness. 1999 Dec;39(4):336-40.

Physical Performance and Function

The physical performance and function exams were completed at the baseline visit, dental visit, sleep visit, visit 2, visit 3,sleep visit 2 and visit 4. Specifically,

Grip Strength	Baseline, Dental, Sleep, Visit 2, Visit 3, Sleep Visit 2, Visit 4
6 Meter Walk	Baseline, Dental, Sleep, Visit 2, Visit 3, Sleep Visit 2, Visit 4
Narrow Walk	Baseline, Dental, Sleep, Visit 2, Visit 3, Sleep Visit 2
400 Meter Walk	Visit 4
Chair Stands	Baseline, Dental, Sleep, Visit 2, Visit 3, Sleep Visit 2, Visit 4
Nottingham Power Rig	Baseline, Dental (Portland site only), Visit 2, Visit 3, Sleep Visit 2 (Portland site only)
SPPB Balance Test	Visit 4

In general, investigators/analysts may want to consider those who were unable (due to fatigue or physical reasons, NOT because of machine/equipment failure) to do the tests in their models. For example, instead of examining the narrow walk pace as a continuous variable, NFNWPACE might be divided into quintiles, with those who were unable to achieve a valid time (NFNWABLE=0) categorized in the slowest (worst) quintile. Another option would be to use the NFNWNUM variable, which categorizes the men according to the number of narrow walk trials they were able to complete (none, 1, or 2 trials). Generally, participants should be excluded when a machine failure occurred and not grouped with the “unable” participants.

The change variables between baseline and each post-visit, and between two post-visits were created in the post visit dataset (e.g., change variables between baseline and sleep visit 1 are included in the VS1 dataset).

Several calculated variables were derived from the performance tests and are described below:

Chair Stands

NFSTDARM

This variable indicates if a participant uses his arms to stand up during the timed repeated chair stands. As of the August 2006 data release, participants who could not complete the measurement (because they were unable to stand up one time unassisted) are coded as 1 (yes) rather than .M (not applicable). Please note that for all other repeated chair stand test form variables, men who were not able to do the single chair stand test are still set to .M for these variables.

NFSTAND1

This is the single chair stand variable, and it provides the complete special missing value codes for all the chair stand measures. Note that other chair stand values coded as .M (not applicable) may need to be recoded to the special missing values from NFSTAND1 to be able to appropriately include the "unable" participants in the analyses. .U is used for participants who are physically unable to do the single chair stand; it is recommended that these participants be included in analyses when possible.

NFTIME5A (included as of the February 2011 data release)

This variable is a version of NFTIME5, the number of seconds required to complete the repeated chair stands, which sets men who used their arms some or all of the time set are to .U (unable). Thus, NFTIME5A includes only men who were able to complete the repeated chair stands without using their arms, while NFTIME5 includes all men who completed the repeated chair stands regardless of arm use.

6 Meter Walk & Narrow Walk

NFSTPLGT

This is the average step length for the usual pace and is calculated using the number of steps from both trials required to walk the 6-meter course.

NFWLKSPD

This variable for walking speed is for the usual pace, and is calculated using both times. At Visit 3, some of the men who had a home visit completed the test on a course that was only 2 or 3 meters long. Since people have a slower walking speed on shorter courses, the walking speed for these men were converted to the speed that would have been seen on a 6-meter course had it been available. The conversion formulas were taken from Guralnik 2000 Journal of Gerontology: MEDICAL SCIENCES, 2000, Vol. 55A, No. 4, M221–M231. Investigators/analysts may wish to perform sensitivity analyses excluding those men whose NFWLKSPD values were converted (i.e., those with NFHMLWC with values of 2 or 3.

NFWLKSPA (included as of the February 2011 data release)
This variable is a version of NFWLKSPD which sets men who used aids during either of the 6-meter walk trials to .U (unable).

NF6MWTM
This is the participant's best time for completing the 6-meter usual pace.

NF6MPACE
This is the version of the walking speed measure for usual pace that uses the participant's best time.

NF6MPACA (included as of the February 2011 data release)
This variable is a version of NF6MPACE which sets men who used aids during either of the 6-meter walk trials to .U (unable).

NF6MABLE (included as of the February 2011 data release)
This is a 3-category variable pertaining to the ability to complete the 6-meter walk trials. It has the following values:

0=completed both trials without aids
1=completed both trials but used aids for at least 1 trial
2=unable to complete both trials

At baseline, men who had "not attempted" for at least one of the trials were coded as NF6MABLE=.R (refused). From the Dental visit onward, men who had "no, unable to assess" for at least one of the trials were coded as NF6MABLE=.N (cannot evalulate) since it is unknown if the man refused to do the trial or if the trial was not completed for another reason. If aid use was missing for at least one of the trials, NF6MABLE was set to .A (missing).

NFNWTIME
This is the participant's best time for completing the narrow walk.

NFNWPACE
This is the walking speed for the narrow walk. It is calculated using the participant's best time.

NFNWNUM
This variable was added as of the August 2006 data release, and it categorizes the men according to the number of narrow walk trials they were able to complete irrespective of aid use (none, 1 trial, 2 trials). This variable is useful since it allows men who were not able to complete any of the narrow walk trials to be included in analyses. As of the February 2011 data release, men who had 3 or more deviations/unable to complete for either of the first two trials but whose third trial was missing or "not

attempted" had their NFNWNUM set to .A (missing) rather than 0 or 1. This is a protocol violation since the men did not have a chance to do the third trial. As of the August 2011 data release, men who had "trial not attempted" for all trials had NFNWNUM set to .A rather than 0.

NFNWABLE (included as of the February 2011 data release)
This is an indicator for the ability to complete at least one narrow walk successfully, irrespective of aid use. The following description may be used to describe the NFNWABLE variable in a manuscript:

"Men were defined as being able to complete at least one narrow walk trial if they stayed within the lines with 2 or fewer deviations during one or more trials. Men with 3 or more deviations for all three trials, those with 3 or more deviations for one or both of the first two trials and who attempted but were unable to complete the rest of the trials, and those who attempted but were unable to complete any of the trials were identified as unable to complete at least one narrow walk trial. Men who did not attempt any of the trials were set to missing for the narrow walk ability indicator, as well as those who should have completed the third trial due to having 3 or more deviations for at least 1 of the first 2 trials but did not attempt the third trial."

NFNWNUMA (included as of the February 2011 data release)
This variable categorizes the men according to the number of narrow walk trials they were able to complete without the use of aids (none, 1 trial, 2 trials). If aid use for at least one of the trials was missing, then NFNWNUMA was set to .A (missing).

NFNWABLA (included as of the February 2011 data release)
This is an indicator for the ability to complete at least one narrow walk successfully without the use of aids.

NFPCTDIF
This is the difference between the usual pace walking speed and the narrow walking speed using the best time, and is expressed as a percentage of the usual pace walking speed. Participants who were unable to complete both measurements are coded as .M (not applicable). Note that the difference is between the walking speed variables (not the walking time variables).

NFDFSCOR
This represents the categorization of the percent difference between usual and narrow walk speed, derived from NFPCTDIF (>15%, 5-15%, -5 to 5%, <-5%). As of the February 2011 data release, this variable was no longer included in the MrOS data. As an alternative to this variable, it is

suggested that the NFPCTDIF be categorized into quartiles or quintiles as needed.

NFUPSCOR and NFNWSCOR
These variables are no longer included in the MrOS data, as the EPESE (Established Populations for the Epidemiologic Study of the Elderly) categorization for walking speeds are not applicable to the MrOS protocols. As an alternative to these variables, it is suggested that the walking speed and narrow walking speed be categorized into quartiles or quintiles as needed.

400 Meter Walk

The Pittsburgh site needed to use an alternative course for the 400 meter walk that was more than 11 laps. They were unable to record their split times because of this. There was also an issue with the length of the course. At the beginning of the visit, the course length was miscalculated and was only 376 meters long instead of 400 meters. The course was corrected to 400 meters for visits starting 9/2/14. For participants that completed the 376m length course, their values have been scaled up as if they walked the full 400 meters. These participants are denoted with the variable NF4FLAG=1

A reference for the 400m Walk:

Simonsick E, Montgomery P, Newman A et al. Measuring fitness in healthy older adults: The Health ABC Long Distance Corridor Walk. J Am Geriatr Soc 2001;49:1544–1548.

Lange-Maia BS, Newman AB, Strotmeyer ES, Harris TB, Caserotti P, Glynn NW. Performance on fast- and usualpaced 400-m walk tests in older adults: are they comparable? Aging Clinical and Experimental Research. 2015 Jun; 27 (3):309-14. doi: 10.1007/s40520-014-0287-y. PMCID: PMC4422783.

Grip Strength

Grip strength was measured using Jamar dynamometers (Sammons Preston Rolyan, Bolingbrook, IL, USA).

The grip strength variables also have a more informative use of the .U code. .U is used for participants who are physically unable to do the measure; it is recommended that these participants be included in analyses when possible.

GSFLAGEX indicates whether or not a participant met the exclusion criteria for the Grip Strength exam. Men were allowed to complete the grip strength test even if he met the exclusion criteria because the risk of

completing the test was low. Investigators may wish to use this variable to run sensitivity analyses.

GSUNABLE (included as of the February 2011 data release)
This is an indicator for the inability to complete the grip strength trials.

Reference:
Harkonen R, Harju R, Alaranta H. Accuracy of the Jamar Dynamometer. *Journal of Hand Therapy.* 1993; Oct-Dec:259-262.

Nottingham Power Rig

Nottingham Power Rig was measured using Nottingham Power Rig (Nottingham University, Nottingham, UK).

For the Nottingham Power Rig, all missing values for these variables are coded to .N in the baseline and dental dataset. As of the August 2006 data release, 3 variables are included in the baseline and dental datasets to indicate the reason for missing data for the right leg (NPREASR), left leg (NPREASL), and both legs (NPREASB). These variables have the following categories:

0=Has a NP Max value
1=Attempted but Unable
2=Missing because of Possible Physical Limitation
3=Refused, unknown reason
4=Did not attempt, unknown reason (eg. machine was broken)

These variables help separate out the reasons for missingness so that men who were not able to perform the test can be included in analyses.

The data collection forms for the Nottingham Power Rig were updated at Visit 2 to include more informative information regarding the reason the test wasn't completed. Analysts can use the NPRGTBR and NPLFTBR variables to determine if the reason that the data were missing. The variables have the following categories:

1=Machine failure
2=Participant Refused
3=Unable due to a physical limitation

From visit 2 onwards, missing values were coded to .A if the Nottingham power rig measurement was attempted but no measurements were recorded or to .M if the measurement was not attempted due to some other reason as stated above.

From the dental visit onwards the following variables were added.
1. NPBTHBR indicates the reason the test was not completed for either leg, based on the reasons provided for the right and left legs.

2. indicators for the ability to complete the Nottingham trials on the right leg (NPABLER)
3. left leg (NPABLEL),
4. either leg (NPABLEB)
At the dental visit, men who attempted but were unable to do the test and men whose data were missing due to a possible physical limitation were set to “no” for these variables; at Visits 2 and 3 and at the sleep visit 2, men who were unable to do the trial due to a physical limitation were set to “no” for these variables. Participants who refused to do the test were set to .R (refused). At the dental visit, men who did not attempt the test due to unknown reason were set to .A (missing). At Visits 2 and 3 and sleep visit 2, men with machine failure were set to .N (cannot evaluate).

Please note that at the baseline visit, 9 trials were completed on each leg. At the dental visit and at the Sleep visit 2, Portland was the only site to complete the Nottingham. All participants from the other sites were set to .M (not applicable) at these 2 visits. Nine trials were completed on each leg at the Dental visit and 5 trials at the sleep visit 2. At Visit 2 and Visit 3 all sites completed the Nottingham and only 5 trials were completed on each leg.

References for the Nottingham include:

1. Bassey EJ, Short AH. A new method for measuring power output in a single leg extension: feasibility, reliability and validity. *Eur J Appl Physiol Occup Physiol.* 1990;60(5):385-390.

2. Bassey EJ, Fiatarone MA, O'Neill EF, Kelly M, Evans WJ, Lipsitz LA. Leg extensor power and functional performance in very old men and women. *Clin Sci (Lond).* Mar 1992;82(3):321-327.

3. Blackwell T, Cawthon PM, Marshall LM, Brand R. Consistency of Leg Extension Power Assessments in Older Men: The Osteoporotic Fractures in Men (MrOS) Study. *Am J Phys Med Rehabil.* 2009 Nov; 88(11):934-40.

Short Physical Performance Battery (SPPB) Balance Test

The balance tests consist of a side by side stand, semi-tandem stand, tandem stand and one leg stand. Number of seconds held is what is measured.

A reference for the balance test is :

Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB. A short physical performance battery assessing lower extremity function: association with self-reported disability

and prediction of mortality and nursing home admission. J Gerontol. 1994 Mar;49(2):M85-94.

Prostate Health
Prostate health questions were asked at the baseline visit, interim questionnaire, visit 2 and visit 3.

The variable PSSCORE is the AUA Prostate Symptom Score. The following reference can be used:

Barry MJ, Fowler FJ, O'Leary MP, Bruskewitz RC, Holtgrewe HL, Mebust WK, Cockett ATK and the Measurement Committee of the American Urological Association. The American Urological Association symptom index for benign prostatic hyperplasia. Journal of Urology. 1992. 148: 1549-1557.

Sleep General Questions
General sleep questions were asked at the interim questionnaire, sleep visit and sleep visit 2.

Please note, the variables SLINSOM and SLRESTL from the interim questionnaire are yes/no variables that ask if "has a doctor ever told you that you have…". At the Sleep visit and Sleep Visti 2 similar questions were asked as subquestions of "Has a doctor ever told you you have a sleep disorder"(SLSLPDIS), if yes which one. (SLINSOMN, SLRESTLG). They are similar, but not identical, so are named slightly different.

On the interim questionnaire there is a categorical question "How many hours of sleep do you usually get each night?" (SLSLPHR). At the sleep visit and sleep visit 2, this question is also asked, but the response is continuous so the variable is named slightly different (SLSLPHRS).

Epworth Sleepiness Scale (ESS)
The Epworth Sleepiness Scale was asked at the sleep visit and sleep visit 2.

This measures daytime sleepiness, and ranges from 0 to 24, with higher scores representing more daytime sleepiness. (EPEPWORT). This scale is a validated instrument, standard in sleep research.

The main reference used for this is:
Johns MW 1991 A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 14(6):540-5.

There is a standard cutpoint of ESS>10 for excessive daytime sleepiness (EPEDS).

The reference for this cutpoint is:
Johns MW. Sensitivity and specificity of the multiple sleep latency test (MSLT), the maintenance of wakefulness test and the epworth sleepiness

scale: failure of the MSLT as a gold standard. J Sleep Res. 2000 Mar;9(1):5-11.

Pittsburgh Sleep Quality Index (PSQI)
The Pittsburgh Sleep Quality Index was asked at the sleep visit, interim questionnaire 2, sleep visit 2 and visit 4.

This measures reported sleep patterns and sleep problems, including sleep quality (PQPSQUAL), sleep latency (PQPLATEN, PQPSLPM4), sleep efficiency (PQPEFFCY**,** PQPEFFIC**)** and daytime dysfunction (PQDAYDYS). The PSQI is a 19-item questionnaire that has been demonstrated to have high internal consistency (0.83), test-retest reliability (0.85) and diagnostic validity. A global sleep quality score derived from the PSQI can be used to index overall quality of sleep over the prior one-week period. Global sleep quality scores are continuous (PQPSQI range 0-21) with high scores reflecting poor sleep quality. A standard cutpoint defining poor sleepers is PQPSQI>5 (PQBADSLP).

The standard references for this are:

Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ 1989 The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Research 28(2):193-213.

Buysse DJ, Reynolds CF, Monk TH, Hoch CC, Yeager AL, Kupfer DJ 1992 Quantification of subjective sleep quality in healthy elderly men and women. Sleep 14(4):331-338.

The Functional Outcomes of Sleep Questionnaire (FOSQ)
The Functional Outcomes of Sleep Questionnaire was asked at the sleep visit.

This is used to measure daytime consequences of sleep problems. The questionnaire is designed to measure functional status in situations that produce sleepiness. The measure has content validity based on 100% agreement by a panel of experts composed of individuals with expertise in sleep disordered breathing, geriatrics and instrument design. There is an overall scale (FOFOSQ) which ranges from 5 to 24, and six subscales, including vigilance (FOVIGIL), intimacy and sexual relationships (not measured), general productivity (FOPRODUC), activity level (FOACTIV), and social outcome (FOSOCIAL). Test-retest reliability for the scale was .91 when administered a week apart. Patients were asked to rate themselves.

The main reference for this is:

Weaver TE, Laizner AM, Evans LK, Maislin G, Chugh DK, Lyon K, Smith PL, Schwartz AR, Redline S, Pack AI, Dinges DF 1997 An instrument to measure functional status outcomes for disorders of excessive sleepiness. Sleep 20(10)**:**835-43.

Insomnia Severity Index
The insomnia severity index was collected at sleep visit 2.

SLISISCR is the Insomnia Severity Index Score with a range of 0 to 28
SLISICAT divides the Insomnia Severity Scores into categories.

The following reference is for the insomnia severity index.

Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med. 2001 Jul;2(4):297-307.

Restless Leg Syndrome Questionnaire
The restless leg syndrome questionnaire was collected at sleep visit 2.

There are two calculated variables for the International Restless Legs Syndrome Study Group Rating Scale:

SLRLSCOR is the International Restless Leg Syndrome (IRLS) Study Group Rating Scale Score with a range of 0-40.
SLRLSCAT divides the IRLS score into severity cateogries

At Sleep visit 2, Participants filled out the restless legs rating scale if they answered yes to lead in questions 2 or 2a (SLRLDES, SLRLRELV)on the sleep history page of the clinic questionnaire. IRLS scoring dictates that the scale shouldn't be filled out unless they answered yes to all lead in questions 2,2a, 2b and 2c. (SLRLDES, SLRLRELV, SLRLREST, SLRLLATR) We have updated our data starting with the **AUG15** release to set scoring variables for the restless legs syndrome scale on page 2 of the clinic questionnaire (variables SLRLSCOR, SLRLSCAT, SLRLDISC, SLRLMOV, SLRLREL, SLRLSLPD, SLRLTIRE, SLRLSYMP, SLRLOFTN, SLRLSEVR, SLRLAFFR, SLRLMOOD) to missing if participants did not answer yes to all 4 of these lead in questions, unless they reported restless legs syndrome in question 4 of the medical history SAQ (variable SLRESTLG). This affected 187 participants who met criteria for questions 2 and 2a, but not 2b and 2c, and did not report restless legs on the SAQ, who had completed the IRLS scale and were subsequently set to missing. If data for these 187 are of interest to you, please contact the coordinating center. Lead in questions for Restless legs were also asked at the sleep visit, but the syndrome scale was not given, so this change does not affect the sleep visit.

The following references are for the restless leg syndrome questionnaire.

International Restless Legs Syndrome Study Group. Validation of the International Restless Legs Syndrome Study Group rating scale for restless legs syndrome. Sleep Med. 2003 Mar;4(2):121-32.

Abetz L, Arbuckle R, Allen RP, Garcia-Borreguero D, Hening W, Walters AS, Mavraki E, Kirsch JM. The reliability, validity and responsiveness of the International Restless Legs Syndrome Study Group rating scale and subscales in a clinical-trial setting. Sleep Med. 2006 Jun;7(4):340-9.

Allen RP, Kushida CA, Atkinson MJ and RLS QoL Consortium. Factor analysis of the International Restless Legs Syndrome Study Group's scale for restless legs severity. Sleep Medicine. 2003; 4:133-135.

Sleep Disordered Breathing:
For analyses that need to account for sleep disordered breathing post sleep visit 1, (CPAP or oxygen therapy use), the coordinating center has a dataset and code that may help. Please contact the coordinating center for further information. (Note to internal analysts: see N:\MrOSData\Sleep Data System\Sleep Disordered Breathing)

Specimen Collection Information

Serum samples, urine samples and whole blood samples were collected at multiple visits in the MrOS study. Please contact the Coordinating Center regarding specimen availability. There are a handful of bookkeeping variables related to specimen collection in the MrOS datasets.

Fasting information
Fasting information was collected at baseline visit, sleep visit, visit 3, sleep visit 2 and visit 4.

The flag variable SCFAST indicates if fasting sample was collected. If the fast hour is less than 8 hours, SCFAST=0:No; if the fast hour is equal to or larger than 8 hours, SCFAST=1:Yes; if the fast hour information was missing, SCFAST=.A: Missing.

The variable SCUFAST indicates if a fasting urine sample was collected.

Sleep Visit Urine Collection Bookkeeping Variables
Sleep visit urine collection information was collected at baseline sleep visit and sleep visit 2.

The flag variable SCUPSG indicates if the spot urine sample was collected on the morning after the PSG data was collected. If the specimen was collected on the morning after the psg, SCUPSG=1:Yes; If it was collected on the different day, or if urine sample was collected, but PSG data was not collected, SCUPSG=0:No; If no urine sample was collected, SCUPSG=.M:Not applicable.

A 24-hour urine sample was collected in participants from the Portland clinic at the first sleep visit. The following variables are included in the sleep visit 1 dataset: SC24URINE (Was 24-hour urine collected?), SC24UVOL (Total volume of the 24-hour urine collection) SC24U24H (Was 24-hour urine collected for full 24 hours), and SC24UALL (Does 24-hour urine sample include all voids). All 24-hour urine variables for clinic sites other than Portland are set to .M:Not applicable.

Tobacco/Alcohol Use

Tobacco and alcohol use data were collected at baseline, interim questionnaire, sleep visit, dental visit, visit 2, visit 3, interim questionnaire 2, sleep visit 2 and visit 4. .

At visit 4, no questions about alcohol use were asked. Questions about currently smoking were part of the activity monitor questionnaire, and was not answered by all participants.

At Sleep Visit 2 questions similar to those asked at baseline regarding length of time smoking and amount smoked in the past are asked as part of the respiratory symptoms questionnaire. The wording of the variables at baseline and Sleep Visit 2 differ slightly, so variables have different names.

TUPACKYR (Baseline), TUPACKY2 (Sleep Visit 2)
This is the main variable for smoking status. Please note that people who have never smoked are coded with the special missing value code of .M:Not applicable.

TUPACKY3 (Sleep Visit 2)
This variable is similar to TUPACKY2, but incorporates more information in the calculation regarding quitting smoking.

TURSMOKE (Baseline, interim questionnaire, sleep visit, dental visit, visit 2, visit 3, interim questionnaire 2, visit 4)
This is the variable that should be used to classify participants as current/ever/never smokers.
At baseline, TURSMOKE is based on the combination of the variables TUSMOKE “Have you smoked at least100 cigarettes (5 packs) in your entire lifetime?” and the variable , “Do you smoke cigarettes now?”.
At the follow-up visits, the TURSMOKE variable is created combining information from the baseline TURSMOKE variable and all prior and current answers to TUSMKNOW. Note that at visit 4, “do you smoke cigarettes now?” is asked on the activity monitor form, so is only filled out by those who received an activity monitor. This included clinic and SAQ ppts.

TURSMOK2 (Sleep Visit 2)
At the Sleep Visit 2, TURSMOK2 is based on the combination of the variables LFSMOKE “Have you smoked cigarettes (no means less than 20 packs of cigarettes or 12 oz. of tobacco in your lifetime or less than 1 cigarette a day for one year at any time in your life)” and the variable LFSMNOW, “Do you smoke cigarettes now (as of one month ago)?”.

TUSMYRST (Baseline), TUSMYRS2 (Sleep Visit 2)
This calculates the number of years the participant has stopped smoking. If the participant is a current smoker or has never smoked the value will be .M.

TUDRPRWK
This calculates the number of drinks per week, on average, that a participant consumes over the 12 months before baseline. Note that individuals who indicated that they had not had at least 12 drinks in the twelve months before baseline (TU12DRIN) are coded as 0.

TUCAGE
This sums the four CAGE questions given the participant answered yes to the question “ever had at least 12 drinks in your entire life” (TUDRINKA). For more information on the CAGE questions and score please refer to:

> Ewing JA. Detecting Alcoholism: The CAGE Questionnaire. Journal of American Medical Association. Volume 252. Pages 1905-1970.1984.

TUSMKNOW

This is the main variable for current smoking status collected at the sleep visit.

There were also some questions at the sleep visit that asked about alcohol and tobacco use on the night of the in-home polysomnography (POXWINE, POXLIQ, POXBEER, POXCIG, POXPIPE, POXCIGAR). A variable summarizing Alcohol use is available as of the FEB19 release, called POXDRNK.

There are a few participants who have somewhat contradictory data for these variables and the “TU” variables. There are 3 participants who say they do not currently smoke cigarettes (TUSMKNOW) but report smoking cigarettes the night of the polysomnography, and 4 participants who report not smoking a pipe or cigar regularly (TUPIPEC) but report smoking a pipe or cigar the night of the polysomnography(POXCIGAR, POXPIPE). There are 9 participants who report that the did not drink at least 12 alcoholic drinks in the past year(TU12DRIN), but report drinking the night of the polysomnography (POXWINE, POXBEER, POXLIQ, POXDRNK).

Vision variables
Vision data was collected at the baseline clinic exam. A self-reported vision question was asked at visit 2, visit 3 and visit 4.

The main variables to use from the Functional Vision exam at baseline are FV2050, FVBLLTRC, FVDISPAR, FVLCS, FVLOGMAR.

FVBLLTRC
This is the corrected visual acuity from the Bailey Lovie test. Again, corrected indicates that the participant wore glasses and/or contacts if necessary for the test. If a participant was unable to read the chart at the standard distance (10 feet) then the test was re-administered at 5 feet. The visual acuity score for these participants was adjusted by subtracting 15 from the score recorded in the raw

data file to account for the alternative distance. This resulted in a very few negative values for those participants with extremely poor vision. The variable FVBLDIST indicates the distance used for the Bailey Lovie test.

FV2050
FV2050 is a calculated variable that indicates whether corrected visual acuity is 20/50 or worse. Please note that we collect corrected acuity measurements, meaning that the participant is wearing glasses and/or contacts if necessary for the eye exams. Again, this is adjusted for the distance used.

FVLOGMAR
Minimum angle of resolution, or MAR, is the reciprocal of Snellen acuity. Poor visual acuity has a minimum angle of resolution exceeding one MAR (e.g., poor visual acuity of 20/200 is equivalent to 10 MAR, or ten times the normal-vision minimum angle). We do not release a MAR variable in the MrOS data set. Instead, the variable FVLOGMAR should be used, which is the base-ten logarithm of the minimum angle of resolution (MAR). The normal vision LogMAR value is zero, while a poor visual acuity LogMAR value exceeds zero (e.g., 20/200 has a LogMAR value of log(200/20) = log(10) = 1). Measured this way, the FVLOGMAR variable is at an interval level (linear) therefore can be analyzed using paramteric statistical techniques.

FVLCS
FVLCS is a calculated variable based on the Pelli-Robson test. A Pelli-Robson score of 2.0 indicates normal contrast sensitivity of 100 percent. Scores less than 2.0 signify poorer contrast sensitivity. Pelli-Robson contrast sensitivity score of less than 1.5 is consistent with visual impairment and a score of less than 1.0 represents in visual disability. This score (1.0) represents an approximately 10-fold loss of contrast sensitivity. That is, a person with contrast sensitivity of 1.0 requires 10 times as much contrast to see as compared with a person with normal vision. As with the acuity measures, the log value is used to enable parametric statistical analyses.

FVDISPAR
This is the disparity variable for depth perception. Higher disparity indicates worse depth perception. More information about this variable will be included in the next data release. For depth perception, the plate thickness variables can also be used.

The following references can be used for the functional vision measures:

Bailey I, Lovie J. New design principles for visual acuity letter charts. Am J Optom Physiol Opt 1976;53:740–5.

Pelli DG, Robson JG, Wilkens AJ. The design of a new letter chart for measuring contrast sensitivity. Clin Vis Sci 1988;2:187–99.

Frisby JP. The Frisby sterotest. Br Orthop J 1980;37:108–12.

Frailty Index information

The Frailty Index, similar to that used by the cardiovascular health study (CHS), has been calculated for all visits where possible (baseline, sleep visit, visit 2, visit 3 and sleep visit 2). These calculated variables are not released because they rely on the distributions of each specific dataset. If you need them please contact the Coordinating Center for both data and information about the calculation and use of the variables.