ANESTHESIOLOGY

Accuracy of Physical Function Questions to Predict Moderate-Vigorous Physical Activity as Measured by Hip Accelerometry

Daniel S. Rubin, M.D., M.S., Megan Huisingh-Scheetz, M.D., Anthony Hung, B.S., R. Parker Ward, M.D., Peter Nagele, M.D., M.Sc., Ross Arena, Ph.D., P.T., Donald Hedeker, Ph.D.

ANESTHESIOLOGY 2019; 131:992–1003

EDITOR'S PERSPECTIVE

What We Already Know about This Topic

- Functional capacity is thought to be an important part of preoperative assessment, but it is hard to assess without formal testing
- Standardized physical function questions might identify patients with adequate capacity (greater than or equal to 4 metabolic equivalents)

What This Article Tells Us That Is New

- Results from standardized physical function questions and hip accelerometers were compared in 522 participants
- Physical function questions were sensitive but nonspecific
- Other approaches to assessing physical functional status should be considered

ABSTRACT

Background: Functional capacity assessment is a core component of current perioperative cardiovascular evaluation and management guidelines for noncardiac surgery. The authors investigated the ability of standardized physical function questions to predict whether participants engaged in moderate physical activity as measured by hip accelerometers.

Methods: Participant responses to physical functioning questions and whether they engaged in moderate physical activity were extracted from the National Health and Nutrition Examination Survey (2003 to 2004 and 2005 to 2006). Physical activity intensity was measured using hip accelerometers. Adult participants with at least one Revised Cardiac Risk Index condition were included in the analysis. Standardized physical function questions were evaluated using a classification and regression tree analysis. Training and test datasets were randomly generated to create and test the analysis.

Results: Five hundred and twenty-two participants were asked the physical functioning questions and 378 of 522 (72.4%) had a bout of moderate-vigorous activity. Classification and regression tree analysis identified a "no difficulty" response to walking up 10 stairs and the ability to walk two to three blocks as the most sensitive questions to predict the presence of a 2-min bout of moderate activity. Participants with positive responses to both questions had a positive likelihood ratio of 3.7 and a posttest probability greater than 90% of a 2-min bout of moderate-vigorous activity. The sensitivity and specificity of positive responses to physical functioning questions in the pruned tree were 0.97 (95% CI, 0.94 to 0.98) and 0.16 (95% CI, 0.10 to 0.23) for training data, and 0.88 (95% CI, 0.75 to 0.96) and 0.10 (95% CI, 0.00 to 0.45) for the test data. Participants with at least one 2-min bout of moderate activity had a greater percentage of overall daily active time (35.4 ± 0.5 *vs.* 26.7 ± 1.2; $P = 0.001$) than those without.

Conclusions: Standardized physical function questions are highly sensitive but poorly specific to identify patients who achieve moderate physical activity. Additional strategies to evaluate functional capacity should be considered.

(ANESTHESIOLOGY 2019; 131:992–1003)

Functional capacity as used in the preoperative assessment may be defined as the ability to perform submaximal physical activities during daily life, and it plays a core role in the current practice guidelines on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery.[1–3] Functional capacity is subjectively assessed by asking patients their ability to perform activities requiring 4 or more metabolic equivalents or using a formalized questionnaire, such as the Duke Activity Status Index. The two physical function questions recommended by textbooks are the ability to walk up a flight of stairs or walk one to two blocks without symptomatic limitations.[4] Patients unable to achieve an activity level of 4 metabolic equivalents are at increased risk of adverse cardiac events.[5–13] Subjective functional capacity assessment thus impacts perioperative risk stratification and dictates whether patients should be considered for further cardiac testing before noncardiac surgery.

However, the ability of standardized individual physical activity questions to accurately assess functional capacity in patients with an increased risk of major adverse cardiac

This article is accompanied by an editorial on p. 960. Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal's Web site (www.anesthesiology.org).

Submitted for publication February 11, 2019. Accepted for publication June 17, 2019. From the Departments of Anesthesia and Critical Care (D.S.R., P.N.) and Public Health Sciences (D.H.), the Sections of Geriatrics and Palliative Medicine (M.H.-S.) and Cardiology (R.P.W.), the Pritzker School of Medicine (A.H.), University of Chicago, Chicago Illinois; and the Department of Physical Therapy, College of Applied Health Sciences, University of Illinois at Chicago, Chicago, Illinois (R.A.).

 Anesthesiology 2019; 131:992–1003. DOI: 10.1097/ALN.0000000000002911

events is unclear. Previous work in nonoperative settings has observed that individuals frequently under- or overestimate their physical activity in response to physical function questions when compared to accelerometer-based physical activity measurements.[14] Self-reported physical activity tools are particularly poor among individuals with low routine activity levels.[15] Additionally, a 2018 multicenter trial found that subjective functional capacity assessment does not accurately predict adverse perioperative outcomes.[16]

Accelerometers offer a novel approach to measuring the intensity and duration of physical activity and validating responses to questions regarding physical function. Accelerometers can detect the presence of moderate-vigorous physical activity, which includes the 4 metabolic equivalents threshold.[14] The National Health and Nutrition Examination Survey is a nationally representative sample of the U.S. population that measured physical activity using a hip-worn accelerometer for 7 days during the 2003 to 2004 and 2005 to 2006 time period, and in addition, asked participants about their physical function.

Our study had two main aims. The first was to use the National Health and Nutrition Examination Survey to establish the accuracy of physical function questions to predict the presence of moderate-vigorous physical activity as assessed using accelerometers during 1 week of wear time. The second was to evaluate for differences in overall physical activity between participants who achieve moderate-vigorous physical activity and those who do not. We hypothesized that any single physical function question would not adequately distinguish between participants who routinely participate in moderate-vigorous physical activity and those who do not. Furthermore, we sought to determine if the presence of at least one bout of moderate-vigorous physical activity would identify significant differences in overall activity between these two groups.

Materials and Methods

National Health and Nutrition Examination Survey

National Health and Nutrition Examination Survey is a complex, multistage probability survey conducted by the National Center for Health Statistics designed to assess the health and nutritional status of the civilian noninstitutionalized U.S. population. The survey design has been detailed elsewhere.[17] Briefly, the survey includes participants of all ages and oversamples several under-studied groups including adults older than 70 yr of age, low-income white persons, and non-Hispanic black and Mexican-American persons. Trained interviewers administered a household interview and performed an on-site examination. Interview data included demographic, dietary, and health-related questions and the examination included medical and physiologic measurements. The National Health and Nutrition Examination Survey cycle for the 2003 to 2004 and 2005 to 2006 included a substudy that measured the daily physical activity of the participants using hip-worn accelerometers. Only data from these two cycles were analyzed for the current study. The University of Chicago institutional review board (Chicago, Illinois) deemed this study exempt since it used a publicly accessible data source (IRB No. 18–1268).

Participants

Participants were included in the analysis if they had at least one Revised Cardiac Risk Index condition. Such participants are considered at increased risk (greater than or equal to 1%) of major adverse cardiac events if they have at least one positive Revised Cardiac Risk Index condition during intrathoracic, intraabdominal, or suprainguinal vascular surgery (high-risk surgery) according to current guidelines.[3] Conditions were identified using questionnaire and laboratory data from the National Health and Nutrition Examination Survey and included diabetes requiring insulin (Diabetes Questionnaire 050: "[Are you] Taking insulin now?"), congestive heart failure (Medical Condition Questionnaire 160B "[Have you] Ever [been] told you have congestive heart failure?"), coronary artery disease (Medical Condition Questionnaire 160C "[Have you] Ever [been] told you have coronary artery disease?"; Medical Condition Questionnaire 160D "[Have you] Ever [been] told you had angina/angina pectoris?"; Medical Condition Questionnaire 160E "[Have you] Ever [been] told you had a heart attack?"), cerebrovascular disease (Medical Condition Questionnaire 160F "[Have you] Ever [been] told you had a stroke?"), and chronic kidney disease (serum creatinine greater than 2.0 mg/dl or Kidney Conditions Question 025 "[Have you] Received dialysis in [the] past 12 months?").[18] The serum creatinine from the 2005 to 2006 National Health and Nutrition Examination Survey cycle was recalibrated according to the analytic notes provided to ensure comparability with standard creatinine values.[19]

Functional Capacity

Questions from the physical functioning questionnaire were used to evaluate self-reported functional capacity. The physical functioning questionnaire (PFQ) was not presented to all participants in the National Health and Nutrition Examination Survey study and these participants were excluded from the current study. Participants under the age of 59 who did not require any special equipment to walk (*e.g.,* walker or cane), had no limitations keeping them from working, or experienced no confusion or memory problems were assumed to have no functional capacity limitations. Physical functioning questions were included in the analysis based on the likelihood that they would be able to discriminate participants based on activity levels they could achieve.

The questions included were: (1) "[Do you have] difficulty walking for a quarter mile/2 to 3 blocks?" (PFQ061B); (2) "[Do you have difficulty] walking up 10 stairs?" (PFQ061C); (3) "[Do you have] lifting or carrying difficulty?" (PFQ061E);

(4) "[Do you have] house chore difficulty?" (PFQ061F); (5) "[Do you have] preparing meals difficulty?" (PFQ061G); and (6) "[Do you have difficulty] walking between rooms on [the] same floor?" (PFQ061H). Responses to the physical function questions were standardized and included: "No difficulty," "Some difficulty," "Much difficulty," "Unable to do," "Do not do this activity," "Refused," "Don't know," or "Missing." In the sample used for the current study, no responses included "Refused" or "Don't know" and all physical function question data were present.[18]

Accelerometry Measures

The National Health and Nutrition Examination Survey used a uniaxial accelerometer (ActiGraph AM-7164; ActiGraph, USA) to measure physical activity among participants older than the age of 6.[20,21] Participants randomly assigned to this substudy were told to wear the monitor at home for 7 consecutive days. Participants who used wheelchairs or had other impairments that prevented walking or wearing the physical activity monitor device were excluded from the substudy. The monitor was placed on an elasticized fabric belt that was worn on the right hip. The device was not water resistant and was removed during bathing/water activities, preventing capture of these activities. Additionally, subjects were instructed to remove the monitor at bedtime and thus, this time period was not captured.

The monitors were programed to start recording at 12:01 AM the day after the participant's health examination. Data from the National Health and Nutrition Examination Survey was obtained using the nhanesdata package and processed using the nhanesaccel package.[22] Activity was summarized into unitless counts per 1-min epoch and processed using the R package nhanesaccel. Non–wear time was defined as any interval 60 min or longer in which all count values were 0. Consistent with previous studies, monitoring days with more than 600 min (more than 10 h) of wear time were considered valid for analysis.[23] At least 4 valid days of wear time were required to be considered a representative characterization of the participant's activity.[23]

The primary outcome measure was the presence of at least one 2-min bout of moderate-vigorous physical activity during the 7-day accelerometer wear time. A 2-min bout was selected because a patient walking at a steady 4-mph pace should take 2 min to travel two blocks. Further, the activity intensity of moderate-vigorous physical activity includes the 4 metabolic equivalent cutoff recommended by the current perioperative practice guidelines.[3,14] Thus, a 2-min bout was deemed the lowest threshold of sustained moderate-vigorous physical activity to meet the guidelines requirements for adequate functional capacity.[3] Activity intensity was identified using predefined cut-points that have been previously validated in the National Health and Nutrition Examination Survey data.[14] Activity intensities were defined as moderate-vigorous (greater than 2,020 counts per min), lifestyle (760 to 2019 counts per min), light (100 to 759 counts per min), and sedentary (less than 100 counts per min).[14] Lifestyle activities include household chores, gardening, and golf and is performed at a lower intensity than moderate-vigorous physical activity.[24]

Secondary outcome measures included average percent of the day spent in sedentary, light, lifestyle, and moderate-vigorous activity which were calculated by summing the minute spent in each activity intensity and dividing by the total number of valid minute for each valid day. We also calculated the average total daily activity counts for each valid day by summing the total activity counts for each day and dividing by the number of wear days. We also calculated mean activity counts during the total wear time. Average daily steps were unavailable for the 2003 to 2004 cycle and so were not included in our analysis.[23] All models were adjusted for total wear time and the number of weekend days worn (0, 1, 2).

Demographic data collected from the participants included age (yr), sex (male or female), race (white, black, Mexican-Hispanic, other Hispanic, other), highest education achieved (less than ninth grade, ninth to eleventh grade, high school diploma/General Education Development Degree, some college or associate's degree, college or above), height (cm), weight (kg), and body mass index (kg/m^2).[25]

Statistical Analysis

The National Health and Nutrition Examination Survey uses a complex, multistage, probability survey design and requires appropriate weights be applied for accurate national population and standard error estimates. For the combined analysis of 2003 to 2004 and 2005 to 2006, we created 4-yr sample weights to account for the different reference populations.[26] Estimates created from this study are thus representative of the U.S. population at the mid-point of the combined survey period (January 1, 2005). Continuous variables are reported as mean ± SD, and categorical variables are reported as frequency. Continuous variables were compared using an adjusted Wald test of the means and categorical variables were compared using Pearson chi-square test. Survey weights were applied to all estimates and statistical tests to ensure accurate standard errors. No *a priori* power calculation was performed and the analysis was based on the available data from the National Health and Nutrition Examination Survey.

Model Development

Classification and regression tree analysis produces a binary decision tree through recursive partitioning of the data to identify variables with the most explanatory power to predict a chosen response variable. Classification and regression tree analysis considers every value of a predictor variable as a potential split point, and the optimal split is chosen such that the resulting two subgroups are more homogenous with respect to predicting the response variable.

Classification trees select the first node through identifying the variable with the most explanatory power, thus producing two intermediate nodes. Intermediate nodes can be further bifurcated until the nodes reach their minimum size or until no significant improvements can be made in the decision tree through splitting. Nodes were prevented from further splitting if they contained less than 10% of the sample or the complexity parameter threshold (10^{-6}) was reached. The complexity parameter represents the tradeoff of how well the tree explains the data and the overall complexity of the tree (number of nodes), thus creating a tree that maximizes prediction and minimizes complexity and overfitting of the training data. The accuracy of the pruned tree was compared to that of a full tree constructed without any limitation on node sizes to ensure no significant loss of predictive ability when using the pruned tree.

The unweighted sample was 522 of 15,915 (3.3%) of the entire National Health and Nutrition Examination Survey sample and was randomly split into a training (90%, n = 469) and validation data set (10%, n = 53). The training sample consisted of 90% of the observations to maximize the accuracy of the classification tree. To offset the risk of overfitting the model we used stringent pruning criteria to avoid deep node splits in the data which would have a greater tendency to overfit our training data. Classification and regression tree analysis was performed to predict the presence of a 2-min bout of moderate-vigorous physical activity using physical function questions, as implemented using the R packages rpart and randomForest.[27,28] Nodes in the classification tree were restricted from further splitting if the complexity parameter threshold was reached and the depth of the tree was limited to two steps. The random forest model is an ensemble classification method that involves the construction of multiple bootstrapped classification trees. Plots from a random forest generated from the data were used to identify the most important predictors using the Gini impurity criterion. Weights were not applied in the classification and regression tree analysis as the analysis focused on classification of the data rather than statistical inference.

Sensitivity Analysis

A *post hoc* sensitivity analysis was performed upon an expanded sample size. Patients with a diagnosis of diabetes (Diabetes Questionnaire 010: "[Has a] Doctor told you [that you] have diabetes?") were included in the sensitivity analysis even if they reported they were not currently taking insulin as validation studies of the Revised Cardiac Risk Index have demonstrated the inclusion of only diabetic patients receiving insulin does not improve the model.[19] The unweighted sample was 695 of 15,915 (4.4%) of the entire National Health and Nutrition Examination Survey sample and was randomly split into a training (89%, n = 625) and validation data set (11%, n = 70). The tables and figures for the sensitivity analysis can be seen in the online supplemental index.

We performed additional classification and regression tree analyses on longer bouts of moderate-vigorous physical activity of 4 min and 6 min in length; however, we do not report these results. The percentage of participants who engaged in moderate-vigorous activity of 4 min (195 of 522, 37.4%) and 6 min (142 of 522, 27.2%) in the original sample were low enough that the analysis could not improve node classification through the use of physical function questions.

Statistical analyses were carried out using STATA-MP V14 (Statacorp, USA) and R V3.5.1 (http://www.r-project.org). All statistical analysis, except for the classification and regression tree analysis, accounted for the complex survey design of the National Health and Nutrition Examination Survey, and survey weights were adjusted for the two cycles of data. Survey analysis was done using the svy and subpop commands of STATA.

Results

A total of 852 (age 20 yr or older) unweighted participants in the survey had at least one Revised Cardiac Risk Index condition. Five hundred and twenty-two were asked the physical functioning questions, all of whom had at least 4 valid days of accelerometer wear time. This cohort represents an estimated 10,174,803 persons in the United States. The mean age for all participants who were asked the physical functioning questions was 69 ± 11 yr and 56% were male. The Revised Cardiac Risk Index condition with the greatest prevalence was coronary artery disease (66.8%), followed by cerebrovascular disease (23.0%), congestive heart failure (24.6%), diabetes requiring insulin (15.8%), and chronic kidney disease (2.7%; table 1). Participants with two or more Revised Cardiac Risk Index conditions totaled 26.8%.

At least one 2-min bout of moderate-vigorous physical activity was present in 72.4% of participants during accelerometer wear time (table 1). Participants with one 2-min bout of moderate-vigorous physical activity were younger and more likely to be male. Participants without a bout of 2 min of moderate-vigorous physical activity were more likely to have a diagnosis of congestive heart failure and cerebrovascular disease. The two groups did not differ in regard to race, education, or body mass index.

The accelerometer parameters measured throughout the week of physical activity monitoring are listed in table 2. Participants with at least one 2-min bout of moderate-vigorous physical activity had more valid days, more overall minutes of valid wear time and a higher average daily wear time of the accelerometer. Additionally, participants with one 2-min bout of moderate-vigorous physical activity spent a higher proportion of time in light and lifestyle activity than those who did not have a 2-min bout of moderate-vigorous physical activity when controlling for number of valid days and wear-time minutes. Participants without a 2-min bout of moderate-vigorous physical activity had a higher proportion of sedentary time.

Table 1. Participant Characteristics Stratified by a 2-Min Bout of Moderate-Vigorous Physical Activity during the Week of Physical Activity Monitoring

Participant Characteristics	2-Min Bout MVPA, 72.4% (n = 378)	No Bout of MVPA, 27.6% (n = 144)	*P* Value
Age, mean ± SD (yr)	68 ± 11	74 ± 10	0.001
Sex (%)			0.001
Male	62	39	–
Female	38	61	–
Race, (%)			0.582
White	81.8	85.6	–
Black	7.5	8.2	–
Mexican Hispanic	3.8	2.4	–
Other Hispanic	1.4	1.1	–
Other race	5.5	2.7	–
Interview language (%)			0.035
English	98.2	99.8	–
Spanish	1.8	0.2	–
Education, estimate (%)			0.346
< 9th Grade	11.6	8.8	–
9th–11th grade	15.9	17.6	–
High school diploma/GED	27.3	34.2	–
Some college or associate degree	27.3	28.2	–
College or above	17.9	10.8	–
Refused	0	0.5	–
Body Measures, mean ± SD			
Weight (kg)	82.1 ± 16.3	81.2 ± 17.8	0.601
Height (cm)	168.6 ± 9.4	164.5 ± 9.8	0.001
BMI (kg/m^2)	28.8 ± 5.0	29.9 ± 5.5	0.110
RCRI conditions (%)			
Diabetes requiring insulin	14.4	19.4	0.263
Chronic kidney disease	2.5	3.4	0.585
Congestive heart failure	19.9	33.1	0.045
Coronary artery disease	67.2	65.9	0.781
Cerebrovascular disease	21.4	33.0	0.050

Sample size (n) is unweighted. Mean and prevalence are weighted to account for the survey design.

BMI, body mass index; GED, General Education Development degree; MVPA, moderate-vigorous physical activity; RCRI, Revised Cardiac Risk Index.

The responses to the physical function questions stratified by 2-min of moderate-vigorous physical activity are listed in table 3. Participants with a 2-min bout of moderate-vigorous physical activity were more likely to report "no difficulty" walking a quarter mile (two to three blocks), walking up 10 stairs, and performing house chores. The groups did not differ with respect to questions about difficulty lifting or carrying 10 pounds, preparing meals, or walking between rooms on the same floor.

Figure 1 presents the results of the classification and regression tree analysis. The classification tree identified self-reported difficulty walking up 10 stairs and walking two to three blocks as predictive of the presence of a 2-min bout of moderate-vigorous physical activity. The stair walking question with a response of "No difficulty" was the first split in the tree, indicating that this question was the strongest predictor of an adequate functional capacity. Among participants responding that they had "no difficulty" walking up 10 stairs, 78.9% achieved a bout of moderate-vigorous physical activity. The split on the ability to walk two to three blocks was dependent on being able to walk up 10 stairs with "No difficulty." Among participants who reported "No difficulty" walking up stairs and responded they were able to walk two to three blocks, 80.5% achieved a bout of moderate-vigorous physical activity. Participants who responded "unable to do" to walking two to three blocks were classified as unlikely to have a 2-min bout of moderate-vigorous physical activity. Of the participants who responded they had "some" or "much difficulty" walking up 10 stairs, 54.6% participated in a bout of moderate-vigorous physical activity. Of the participants who responded they were "unable" to walk upstairs, 35.0% participated in a bout of moderate-vigorous physical activity. Other physical function questions were not included in the classification tree as they did not improve the performance of the decision tree without adding additional complexity.

Model parameters from the classification and regression tree analysis for the training and test samples are presented in table 4. The sensitivity and specificity of the pruned tree on the training data were 0.97 (95% CI, 0.94 to 0.98) and 0.16 (95% CI, 0.10 to 0.23), respectively. The sensitivity and specificity of the pruned tree on the test data were 0.88

Table 2. Physical Function Questions Stratified by Presence or Absence of a 2-Min Bout of Moderate-Vigorous Physical Activity

Physical Function Question	2-Min Bout MVPA (n = 378)	No Bout of MVPA (n = 144)	*P* Value
Walking for a quarter mile difficulty (%)			0.001
No difficulty	69.8 (264)	45.8 (65)	–
Some difficulty	21.0 (74)	24.5 (38)	–
Much difficulty	4.9 (23)	13.9 (18)	–
Unable to do	2.4 (8)	12.2 (16)	–
Do not do this activity	2.0 (9)	3.6 (7)	–
Walking up 10 steps difficulty (%)			0.007
No difficulty	78.9 (302)	56.4 (80)	–
Some difficulty	13.6 (45)	27.0 (38)	–
Much difficulty	4.7 (20)	8.6 (12)	–
Unable to do	1.5 (6)	5.0 (9)	–
Do not do this activity	1.3 (5)	3.1 (5)	–
Lifting or carrying difficulty (%)			0.085
No difficulty	83.0 (304)	70.6 (103)	–
Some difficulty	9.3 (43)	16.5 (25)	–
Much difficulty	2.9 (11)	37.2 (4)	–
Unable to do	3.2 (17)	7.5 (9)	–
Do not do this activity	0.9 (2)	1.6 (3)	–
Don't know	0.7 (1)	0 (0)	–
House chore difficulty (%)			0.003
No difficulty	75.9 (284)	55.5 (85)	–
Some difficulty	14.7 (57)	29.7 (42)	–
Much difficulty	2.8 (11)	3.7 (4)	–
Unable to do	1.5 (5)	5.5 (5)	–
Do not do this activity	5.0 (21)	5.7 (8)	–
Preparing meals difficulty (%)			0.366
No difficulty	88.8 (329)	82.7 (115)	–
Some difficulty	5.4 (21)	8.4 (13)	–
Much difficulty	0.6 (3)	0.7 (2)	–
Unable to do	0.03 (1)	0.5 (1)	–
Do not do this activity	5.1 (24)	7.7 (13)	–
Walking between rooms on the same floor (%)			0.101
No difficulty	96.2 (362)	93.2 (134)	–
Some difficulty	3.7 (15)	6.7 (9)	–
Much difficulty	0.08 (1)	0.1 (1)	–
Unable to do	0 (0)	0 (0)	–
Do not do this activity	0 (0)	0 (0)	–

MVPA, moderate-vigorous physical activity.

Table 3. Accelerometer Measurements during the Week of Physical Activity Monitoring Stratified by a 2-Min Bout of Moderate-Vigorous Physical Activity

Accelerometer Parameter (Mean ± SD)	2-Min Bout MVPA (n = 378)	No 2-Min Bout (n = 144)	*P* Value
Valid days	6.3 ± 0.9	6.1 ± 1.0	0.018
Valid min of wear time	5401.3 ± 985.3	4972.5 ± 1085.3	0.006
Average daily wear time	848.0 ± 77.1	818.2 ± 86.7	0.002
Average counts per minute	218.7 ± 97.4	110.6 ± 49.4	0.001
Percent time sedentary	64.6 ± 9.9	73.3 ± 10.0	0.001
Percent time active	35.4 ± 9.9	26.7 ± 10.0	0.001
Percent time light activity	27.0 ± 7.0	24.0 ± 8.5	0.013
Percent time lifestyle activity	6.9 ± 4.0	2.6 ± 1.9	0.001
Percent time MVPA	1.4 ± 1.6	0.1 ± 0.1	0.001

Sample size (n) is unweighted. Mean and prevalence are weighted to account for the survey design. Comparisons were adjusted to control for differences in total wear time and number of weekend days worn (0,1,2). A valid day consisted of at least 10 h of activity. Accelerometer cut points to classify activity intensity were: MVPA (more than 2020 counts per min), lifestyle (760–2019 counts per min), light (100–759 counts per min), and sedentary (less than 100 counts per min).

MVPA, moderate-vigorous physical activity.

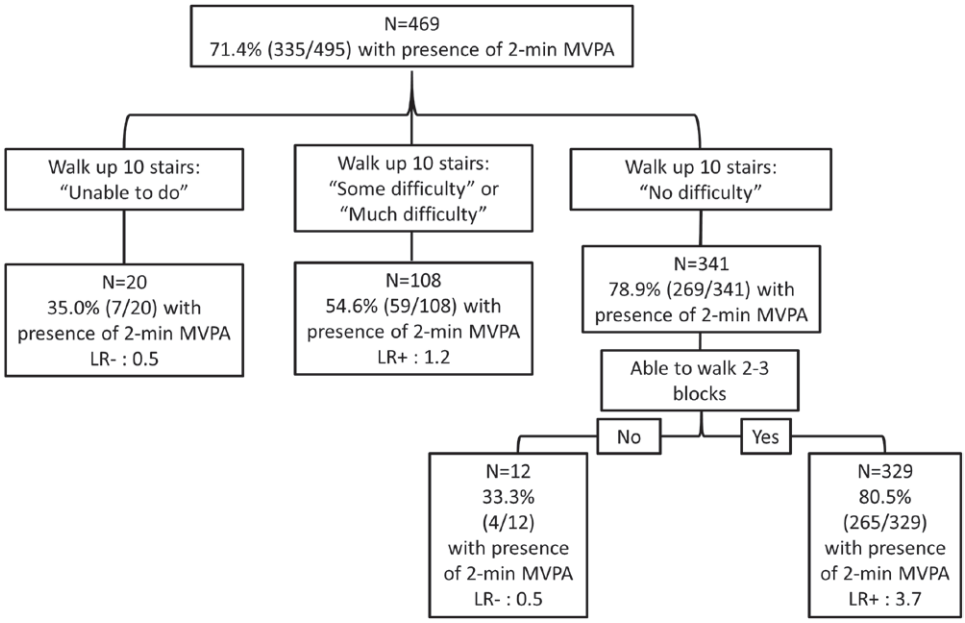

Fig. 1. Classification tree of National Health and Nutrition Examination Survey physical function questions using the presence of a 2-min bout of moderate-vigorous physical activity as the outcome variable. The first split is for participants who responded they had no difficulty walking up 10 steps, of which 78.9% had a 2-min bout of moderate-vigorous physical activity. The sensitivity of detecting a 2-min bout of activity increased further when participants answered they had no difficulty walking up 10 stairs and were able to walk two to three blocks. The classification tree was unable to further improve prediction of the outcome variable with additional questions for participants who answered they had some or much difficulty or unable to walk upstairs. Questions: (Stairs) "By yourself and without using any special equipment, how much difficulty do you have walking up 10 steps without resting?" (Walk two to three blocks) "By yourself and without using any special equipment, how much difficulty do you have walking for a quarter mile [that is about 2 to 3 blocks]?" Possible responses to the physical function questions: (1) No difficulty; (2) Some difficulty; (3) Much difficulty; (4) Unable to do; or (5) Do not do this activity. LR+, positive likelihood ratio; LR–, negative likelihood ratio; MVPA, moderate-vigorous physical activity.

Table 4. Model Parameters from Classification and Regression Tree Analysis of Physical Functioning Questions

Model Parameter	Training Data (n = 469)	Test Data (n = 53)
Prevalence of 2-min bout of MVPA	0.71	0.81
Sensitivity	0.97 (0.94–0.98)	0.88 (0.75–0.96)
Specificity	0.16 (0.10–0.23)	0.10 (0.00–0.45)
Positive likelihood ratio	1.15 (1.06–1.24)	0.98 (0.78–1.24)
Negative likelihood ratio	0.21 (0.10–0.42)	1.16 (0.15–8.89)
Accuracy	0.74 (0.69–0.78)	0.74 (0.60–0.85)

Sample size (n) is unweighted. Weights were not applied to the classification and regression tree analysis as the analysis focused on classification of the data rather than statistical inference. The training and test data were randomly partitioned from the full dataset.

MVPA, moderate-vigorous physical activity.

(95% CI, 0.75 to 0.96) and 0.10 (95% CI, 0.00 to 0.45), respectively. Random forest analyses results are presented in figure 2 and support the classification and regression tree analysis that walking up 10 stairs and walking two to three blocks are the strongest predictors of a 2-min bout of moderate-vigorous physical activity as assessed by the mean decrease in accuracy and mean decrease in Gini impurity if they were removed from the classification tree. None of the other physical function questions or Revised Cardiac Risk Index conditions were reliable predictors of the outcome.

The results of the sensitivity analysis can be seen in the Supplemental Digital Content, table S1 (http://links.lww.com/ALN/C36), which contains the participant characteristics and Supplemental Digital Content, figure S1 (http://links.lww.com/ALN/C39), which contains the classification tree. The accelerometer parameters measured throughout the week of physical activity monitoring are listed in Supplemental Digital Content, table S2 (http://links.lww.com/ALN/C37), and demonstrate increased physical activity in participants who engage in a 2-min bout of moderate-vigorous physical activity. Similar to the primary analysis, stair walking is the optimal question identified by the classification and regression tree analysis to improve the prediction of a bout of moderate-vigorous physical activity with splits identified at "no difficulty," "some/much difficulty," and "unable to do." The likelihood ratio for participants who responded, "no difficulty," is 4.2, which is

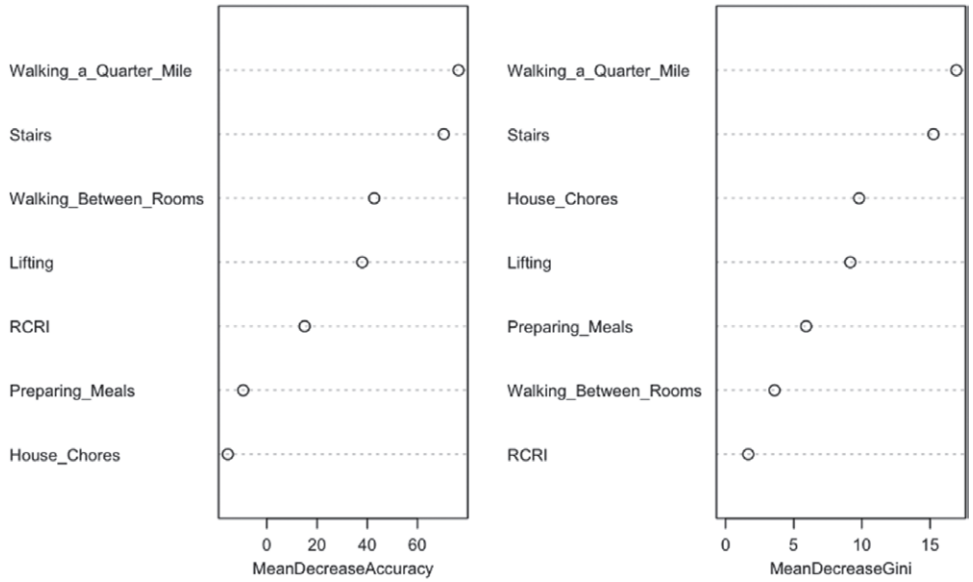

Fig. 2. Results of random forest analyses illustrating the importance of each physical function question in the classification and regression tree analysis. Stair climbing and walking two to three blocks are the most important questions for predicting the presence of a 2-min bout of moderate-vigorous physical activity. (*Left*) Indicates the mean decrease in accuracy of the tree as the result of removing a physical functioning question. (*Right*) The Gini coefficient is a measurement of the likelihood of an incorrect classification of the outcome. Physical function questions that decrease the mean Gini coefficient improve the prediction of that outcome. In each figure, the greater the decrease in accuracy and the greater the mean decrease in Gini impurity is associated with increased variable importance to the final classification and regression tree analysis. RCRI, Revised Cardiac Risk Index.

equivalent to a posttest probability of greater than 90%.[28] The full model parameters are similar to the primary analysis and can be seen in Supplemental Digital Content, table S3 (http://links.lww.com/ALN/C38), and Supplemental Digital Content, figure S2 (http://links.lww.com/ALN/C40), which contains the random forest analyses.

Discussion

In our analysis of the National Health and Nutrition Examination Survey database, self-reported ability to walk up 10 stairs without difficulty best predicted the presence of a 2-min bout of moderate-vigorous physical activity during a week of accelerometer wear among adult participants with at least one Revised Cardiac Risk Index condition. The prediction was further improved when participants responded that they were able to walk two to three blocks, with a positive likelihood ratio of 3.7 and a posttest probability greater than 90% that the participant engaged in a 2-min bout of moderate-vigorous physical activity.[29] Additionally, the sensitivity analysis participants who responded "No difficulty" to walking up 10 stairs also had a posttest probability greater than 90% of a 2-min bout. However, the overall specificity was low in both the training and test data sets, thus calling into question the utility of subjective functional capacity assessment in patients at increased risk of major adverse cardiac events before noncardiac surgery. With respect to our second aim, participants who engaged in a 2-min bout of moderate-vigorous physical activity had a greater percentage of overall active time and a lower percentage of sedentary time as compared to participants who did not engage in a 2-min bout of moderate-vigorous physical activity.

Our data suggest that individual physical activity questions are insufficiently specific to identify patients who do not engage in moderate activity. Functional capacity as assessed by cardiopulmonary exercise testing has been used for decades for perioperative risk stratification and guidelines on the use of cardiopulmonary exercise testing before surgery were published in 2018.[6–9,13,30] Older *et al.* identified an oxygen consumption at ventilatory threshold of less than 11.0 ml · kg^{-1} · min^{-1} as a critical value identifying patients at risk for developing severe morbidity and mortality during major abdominal surgery.[6] In 2016, the American Heart Association emphasized the role of formal cardiopulmonary testing to identify patients at high perioperative risk who are scheduled for an elevated risk surgical procedure, and in whom functional capacity is unknown.[31] However, cardiopulmonary exercise testing is constrained by time, personnel, and cost. In contrast, subjective assessment is easy to perform, and the questions used are surprisingly consistent over time. In a 1999 single center study, Reilly *et al.* used patient self-reported exercise tolerance to predict perioperative complications.[5] Exercise tolerance was

evaluated by asking patients two physical activity questions: (1) "How many blocks can you walk?" and (2) "How many flights of stairs can you climb?"[5] Poor exercise tolerance was defined by a patient's inability to walk four blocks or climb two flights of stairs without symptomatic limitation. Our results confirm that these questions are the most sensitive to identify patients who do not engage in moderate activity. However, current practice may differ from Reilly *et al.* in using different thresholds for a patient to be given a designation of an adequate functional capacity.[3,4,32] The modern approach to functional assessment is highly variable given the type of activities provided in the guidelines that are associated with more than 4 metabolic equivalents. Additionally, the guidelines do not specify a duration of time those activities need to be maintained for, further increasing uncertainty for clinicians performing perioperative risk assessment.[3]

We sought to clarify which, if any, physical activity question most effectively detects the ability to sustain a short duration of moderate-vigorous physical activity. Physical activity questions provide information about the patient's functional capacity and—more broadly—cardiorespiratory fitness. In response to such questions, patients may over- or underestimate their physical abilities when compared to objective measures for a variety of reasons, including poor recall, social desirability, and misinterpretation of the questions.[33] In the National Health and Nutrition Examination Survey database, although 51% of all adults self-reported adherence to national guidelines recommending at least 150 min of moderate-intensity physical activity per week, accelerometer-based measurements revealed that less than 5% were actually adherent.[14] In a large multicenter study by Wijeysundera *et al.,* subjective assessment of functional capacity did not predict cardiac complications in patients undergoing major noncardiac surgery; although subjective assessment was not standardized and conclusions about individual questions cannot be made.[16] In the same study, however, the Duke Activity Status Index did have moderate predictive ability to identify cardiac complications. The Duke Activity Status Index is a structured questionnaire that more formally estimates functional capacity and is also recommended by the guidelines.[3] Our results indicate that the answer to the question, "can you walk up 10 stairs?" has the best sensitivity with respect to an individual's ability to engage in moderate-vigorous physical activity. Further, when participants answered "no difficulty" to walking up 10 stairs and are able to walk two to three blocks, the posttest probability is almost 90% in their having engaged in a 2-min bout of moderate-vigorous physical activity.[29] However, our findings also suggest the overall predictive power of these questions to identify patients who do routinely perform short bouts of moderate-vigorous physical activity is poor.

Preoperative accelerometers or short exercise tests may provide a more objective assessment of functional capacity before noncardiac surgery. In our study, a 2-min bout of moderate-vigorous physical activity during the week of accelerometer wear time differentiated participants by overall activity level. However, it is unclear whether this difference is meaningful for preoperative risk stratification. In previous studies of preoperative accelerometer use, overall accelerometer activity was moderately correlated with cardiopulmonary exercise testing–derived peak oxygen consumption and ventilatory threshold.[34] In addition to accelerometers, short exercise tests may be utilized to identify patients at high risk of complications. Six-minute walk test distance is strongly correlated with the ventilatory threshold and distance can be used to risk stratify patients before major noncardiac surgery.[35,36] Additionally, Reddy *et al.* identified the time to complete an in-clinic stair climb test as the single strongest predictor of perioperative complications in patients undergoing major abdominal surgery, and this test outperformed the American College of Surgeons National Surgical Quality Program score (area under the curve, 0.81 *vs.* 0.62; $P < 0.0001$).[37] More objective measures of patient physical activity, either through patient worn accelerometers or brief exercise tests, may further improve perioperative risk stratification.

Limitations

Our study has several limitations. The National Health and Nutrition Examination Survey database is not a surgical database and may not generalize to surgical patients. Nonetheless, this cohort provides a nationally representative sample of physical activity and responses to physical function questions in participants who would be considered at increased risk of major adverse cardiac events. We did not relate the responses of physical activity questions directly to the results of formal cardiopulmonary exercise testing, the gold standard approach to functional capacity assessment. Rather, our primary outcome was the presence of a continuous 2-min bout of moderate-vigorous physical activity during the week of accelerometer wear time. Because the guidelines specifically emphasize the performance of activities associated with more than 4 metabolic equivalents, we believe our primary outcome is representative of this recommendation and a valid endpoint for analysis. Additionally, our choice of 2 min for the duration of continuous moderate-vigorous physical activity was not explicitly defined in the guidelines, as they do not specify any duration of activity associated with more than 4 metabolic equivalents of work.[3] Thus, we chose the shortest duration (2 min) that would approximate activities included in the practice guidelines (*e.g.,* walking two blocks at a rate of 4 mph). Our choice of accelerometer cut points may have been too conservative. However, we chose cut points that traditionally applied to the National Health and Nutrition Examination Survey dataset to classify moderate-vigorous physical activity. It is possible that this approach may have led to more participants classified as having a poor

functional capacity even though they may have met other criteria.[14,23] In our analysis, 27% of participants were classified as having an inadequate functional capacity, which is consistent with the distribution of previous studies looking at preoperative functional capacity assessment in this patient population.[6,16] Finally, not all of the participants included in the analysis completed all 7 days of the accelerometer protocol, which may introduce bias of being classified as having a poor functional capacity. The bias may have been further amplified by the use of our outcome variable as the presence or absence of a minimum threshold (2-min bout of moderate-vigorous physical activity); however, we did control for the number of weekend days and number of valid days as this can impact the overall results.[38]

Conclusions

In conclusion, we observed in a large nationally representative database that a standardized physical function question focused on walking up stairs was the most sensitive in identifying the presence of a 2-min bout of moderate-vigorous physical activity in participants with at least one Revised Cardiac Risk Index condition. Despite a high sensitivity, however, this single question remains insufficiently specific to identify patients with a poor functional capacity. Given the results of our study, future perioperative cardiovascular evaluation guidelines should consider recalibrating the method and role of subjective functional assessment for risk stratification. Accelerometers and brief exercise tests may improve preoperative risk stratification, but more research is needed to clarify their role.

Acknowledgments

The authors are grateful for the assistance of Sarah Urbut, Ph.D. (Pritzker School of Medicine, University of Chicago, Chicago, Illinois), for her work on the National Health and Nutrition Examination Survey dataset using the R software platform. Additionally, the authors are grateful for the assistance of Avery Tung, M.D., (Department of Anesthesia and Critical Care, University of Chicago, Chicago, Illinois) for his work on editing the manuscript.

Research Support

Mr. Hung is supported by the University of Chicago Medical Scientist Training Program (MSTP; Chicago, Illinois; National Institutes of Health [Bethesda, Maryland] MSTP training grant No. T32GM007281).

Competing Interests

Dr. Rubin is the president of DRDR Mobile Health (Chicago, Illinois), which develops healthcare-focused mobile applications. He has not received any money or payments from the company. The other authors declare no competing interests.

Correspondence

Address correspondence to Dr. Rubin: 5841 South Maryland Avenue, MC-4028, Chicago, Illinois 60637. drubin@dacc.uchicago.edu. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

1. Arena R, Myers J, Williams MA, Gulati M, Kligfield P, Balady GJ, Collins E, Fletcher G; American Heart Association Committee on Exercise, Rehabilitation, and Prevention of the Council on Clinical Cardiology; American Heart Association Council on Cardiovascular Nursing: Assessment of functional capacity in clinical and research settings: A scientific statement from the American Heart Association Committee on Exercise, Rehabilitation, and Prevention of the Council on Clinical Cardiology and the Council on Cardiovascular Nursing. Circulation 2007; 116:329–43
2. Ross R, Blair SN, Arena R, Church TS, Després JP, Franklin BA, Haskell WL, Kaminsky LA, Levine BD, Lavie CJ, Myers J, Niebauer J, Sallis R, Sawada SS, Sui X, Wisløff U; American Heart Association Physical Activity Committee of the Council on Lifestyle and Cardiometabolic Health; Council on Clinical Cardiology; Council on Epidemiology and Prevention; Council on Cardiovascular and Stroke Nursing; Council on Functional Genomics and Translational Biology; Stroke Council: Importance of assessing cardiorespiratory fitness in clinical practice: A case for fitness as a clinical vital sign: A scientific statement from the American Heart Association. Circulation 2016; 134:e653–99
3. Fleisher LA, Fleischmann KE, Auerbach AD, Barnason SA, Beckman JA, Bozkurt B, Davila-Roman VG, Gerhard-Herman MD, Holly TA, Kane GC, Marine JE, Nelson MT, Spencer CC, Thompson A, Ting HH, Uretsky BF, Wijeysundera DN; American College of Cardiology; American Heart Association: 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. J Am Coll Cardiol 2014; 64:e77–137
4. Miller R, Erikson L, Fleisher L, Wiener-Kronish J, Cohen N, Young W: Miller's Anesthesia, 8th edition. Philadelphia, Pennsylvania, Elsevier Saunders, 2014., pp. 1094–7
5. Reilly DF, McNeely MJ, Doerner D, Greenberg DL, Staiger TO, Geist MJ, Vedovatti PA, Coffey JE, Mora

MW, Johnson TR, Guray ED, Van Norman GA, Fihn SD: Self-reported exercise tolerance and the risk of serious perioperative complications. Arch Intern Med 1999; 159: 2185–92
6. Older P, Hall A, Hader R: Cardiopulmonary exercise testing as a screening test for perioperative management of major surgery in the elderly. Chest 1999; 116:355–62
7. Epstein SK, Faling LJ, Daly BD, Celli BR: Predicting complications after pulmonary resection. Preoperative exercise testing vs a multifactorial cardiopulmonary risk index. Chest 1993; 104:694–700
8. Epstein SK, Freeman RB, Khayat A, Unterborn JN, Pratt DS, Kaplan MM: Aerobic capacity is associated with 100-day outcome after hepatic transplantation. Liver Transpl 2004; 10:418–24
9. Carlisle J, Swart M: Mid-term survival after abdominal aortic aneurysm surgery predicted by cardiopulmonary exercise testing. Br J Surg 2007; 94:966–9
10. Nugent AM, Riley M, Megarry J, O'Reilly MJ, MacMahon J, Lowry R: Cardiopulmonary exercise testing in the pre-operative assessment of patients for repair of abdominal aortic aneurysm. Ir J Med Sci 1998; 167:238–41
11. Smith TB, Stonell C, Purkayastha S, Paraskevas P: Cardiopulmonary exercise testing as a risk assessment method in non cardio-pulmonary surgery: A systematic review. Anaesthesia 2009; 64:883–93
12. Snowden CP, Prentis JM, Anderson HL, Roberts DR, Randles D, Renton M, Manas DM: Submaximal cardiopulmonary exercise testing predicts complications and hospital length of stay in patients undergoing major elective surgery. Ann Surg 2010; 251:535–41
13. McCullough PA, Gallagher MJ, Dejong AT, Sandberg KR, Trivax JE, Alexander D, Kasturi G, Jafri SM, Krause KR, Chengelis DL, Moy J, Franklin BA: Cardiorespiratory fitness and short-term complications after bariatric surgery. Chest 2006; 130:517–25
14. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M: Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008; 40:181–8
15. Bonnefoy M, Normand S, Pachiaudi C, Lacour JR, Laville M, Kostka T: Simultaneous validation of ten physical activity questionnaires in older men: A doubly labeled water study. J Am Geriatr Soc 2001; 49:28–35
16. Wijeysundera DN, Pearse RM, Shulman MA, Abbott TEF, Torres E, Ambosta A, Croal BL, Granton JT, Thorpe KE, Grocott MPW, Farrington C, Myles PS, Cuthbertson BH; METS study investigators: Assessment of functional capacity before major non-cardiac surgery: An international, prospective cohort study. Lancet 2018; 391:2631–40
17. NHANES Survey Methods and Analytic Guidelines. Available at: https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx. Accessed June 15, 2018.
18. National Health and Nutritional Examination Survey: Questionnaire Data. Available at: https://wwwn.cdc.gov/nchs/nhanes/Search/DataPage.aspx?Component=Questionnaire&CycleBeginYear=2003. Accessed June 20, 2018.
19. Davis C, Tait G, Carroll J, Wijeysundera DN, Beattie WS: The Revised Cardiac Risk Index in the new millennium: A single-centre prospective cohort re-evaluation of the original variables in 9,519 consecutive elective surgical patients. Can J Anaesth 2013; 60:855–63
20. Center for Disease Control and Prevention: National Health and Nutrition Examination Survey Physical Activity Monitor Data Documentation, Codebook, Frequencies. 2005–2006. Available at: https://wwwn.cdc.gov/Nchs/Nhanes/2005–2006/PAXRAW_D.htm. Accessed July 15, 2018.
21. Center for Disease Control and Prevention: National Health and Nutrition Examination Survey Physical Activity Monitor, Data Documentation, Codebook, Frequencies. 2003–2004. Available at: https://wwwn.cdc.gov/Nchs/Nhanes/2003–2004/PAXRAW_C.htm. Accessed July 15, 2018.
22. Van Domelen D, Pittar W, Harris T: nhanesaccel: Process Accelerometer Data from NHANES 2003–2006. Available at: http://r-forge.r-project.org/projects/nhanesaccel. Accessed June 25, 2018.
23. Tudor-Locke C, Camhi SM, Troiano RP: A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003-2006. Prev Chronic Dis 2012; 9:E113
24. Camhi SM, Sisson SB, Johnson WD, Katzmarzyk PT, Tudor-Locke C: Accelerometer-determined lifestyle activities in US adults. J Phys Act Health 2011; 8:382–9
25. National Health and Nutritional Survey: Demographics Data. Available at: https://wwwn.cdc.gov/nchs/nhanes/Search/DataPage.aspx?Component=Demographics&CycleBeginYear=2003. Accessed June 20, 2018.
26. National Health and Nutritional Examination Survey: How to Construct Weights When Combining Survey Cycles. Available at: https://www.cdc.gov/nchs/tutorials/nhanes/SurveyDesign/Weighting/Task2.htm. Accessed August 12, 2018.
27. Therneau T, Atkinson B. rpart: Recursive Partitioning and Regression Trees. R package version 4.1–13. https://CRAN.R-project.org/package=rpart. Accessed August 17, 2018.
28. Liaw A, Wiener M: Classification and Regression by randomForest. R News 2002; 3: 18–22
29. McGee S: Simplifying likelihood ratios. J Gen Intern Med 2002; 17:646–9

30. Levett DZH, Jack S, Swart M, Carlisle J, Wilson J, Snowden C, Riley M, Danjoux G, Ward SA, Older P, Grocott MPW; Perioperative Exercise Testing and Training Society (POETTS): Perioperative cardiopulmonary exercise testing (CPET): Consensus clinical guidelines on indications, organization, conduct, and physiological interpretation. Br J Anaesth 2018; 120:484–500
31. Guazzi M, Arena R, Halle M, Piepoli MF, Myers J, Lavie CJ: 2016 focused update: Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. Circulation 2016; 133:e694–711
32. Melon CC, Eshtiaghi P, Luksun WJ, Wijeysundera DN: Validated questionnaire vs physicians' judgment to estimate preoperative exercise capacity. JAMA Intern Med 2014; 174:1507–8
33. Kowalski K, Rhodes R, Naylor PJ, Tuokko H, MacDonald S: Direct and indirect measurement of physical activity in older adults: A systematic review of the literature. Int J Behav Nutr Phys Act 2012; 9:148
34. Cui HW, Kirby GS, Surmacz K, Hargrove C, Griffiths J, Turney BW: The association of pre-operative home accelerometry with cardiopulmonary exercise variables. Anaesthesia 2018; 73:738–45
35. Sinclair RC, Batterham AM, Davies S, Cawthorn L, Danjoux GR: Validity of the 6 min walk test in prediction of the anaerobic threshold before major non-cardiac surgery. Br J Anaesth 2012; 108:30–5
36. Shulman MA, Cuthbertson BH, Wijeysundera DN, Pearse RM, Thompson B, Torres E, Ambosta A, Wallace S, Farrington C, Myles PS; Measurement of Exercise Tolerance for Surgery Study Investigators: Using the 6-minute walk test to predict disability-free survival after major surgery. Br J Anaesth 2019; 122:111–9
37. Reddy S, Contreras CM, Singletary B, Bradford TM, Waldrop MG, Mims AH, Smedley WA, Swords JA, Wang TN, Heslin MJ: Timed stair climbing Is the single strongest predictor of perioperative complications in patients undergoing abdominal surgery. J Am Coll Surg 2016; 222:559–66
38. Kocherginsky M, Huisingh-Scheetz M, Dale W, Lauderdale DS, Waite L: Measuring physical activity with hip accelerometry among U.S. older adults: How many days are enough? PLoS One 2017; 12:e0170082