Sudano JJ, Huber GM, Perzynski AT, Lewis SA, and Murray PK. Difficulties in administering and evaluating a 3-minute step test modified for a clinical population. J Sport Human Perf 2014;2(3):1-12. DOI: 10.12922/jshp.0046.2014

JSHP
JOURNAL OF SPORT AND HUMAN PERFORMANCE

ORIGINAL RESEARCH OPEN ACCESS

DIFFICULTIES IN ADMINISTERING AND EVALUATING A 3-MINUTE STEP TEST MODIFIED FOR A CLINICAL POPULATION

Sudano JJ[1,2], Huber GM[3], Perzynski AT[1,2], Lewis SA[1,2], and Murray PK[1,4]

[1]*Center for Healthcare Rearch and Policy, Case Western Reserve University at The MetroHealth System, Cleveland, OH*

[2]*Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH*

[3]*Department of Physical Therapy and Human Movement Sciences, The Feinberg School of Medicine, Northwestern University, Chicago, IL*

[4]*Department of Physical Medicine and Rehabilitation, The MetroHealth System and School of Medicine, Case Western Reserve University, Cleveland, OH*

ABSTRACT

Objective: Measurement of aerobic capacity is an important issue in clinical and research settings. We report on the design, use and attrition rate of a 3-minutes step test modified from the Queens College Step test that we hoped to use as an inexpensive and brief indicator of cardiovascular endurance. **Design:** We conducted a cohort survey and epidemiological study. **Setting:** Two academic medical practices and three community clinics in Cleveland Ohio and Chicago Illinois. **Participants:** Study population included a clinical sample of 1,234 White, Black, English-speaking and Spanish-speaking Hispanic adults aged 45-64. **Methods:** After two modifications of the Queens College Step test we used a 10 inch step height and a cadence of 23 steps per minute to represent our 3-minute step test. Subjects were stopped when they a) completed three minutes, b) when heart rates exceeded 80% of maximal predicted heart rate or reported nausea, dizziness or chest pain, or c) if they requested to stop. **Main Outcome Measurements:** Main outcome measures were completion, dropout and stop rates. **Results:** A total of 28% of subjects were able to complete the test, 36% had to be stopped when heart rates were excessive. The remainder either refused to attempt the test (13%) or requested to stop for symptoms of pain or fatigue (23%). Men and non-Hispanic whites were more likely to complete the test. **Conclusion:** Existing fixed cadence step tests that are ungraded are particularly problematic for assessing aerobic capacity/fitness in a clinical population because of high dropout rates during testing. Further exploration of self-paced tests should focus on the ability of subjects to complete the tests and the validity of those tests to predict aerobic capacity.

Keywords: performance testing, aerobic capacity, step-test, clinical population

INTRODUCTION

The measurement of aerobic capacity is often necessary in both clinical and research settings. At a clinical level, sedentary lifestyles are associated with increases in morbidity and mortality and higher levels of physical fitness are associated with better health. Because of these relationships Centers for Disease Control and other major health organizations have recommended that adults participate in 30 minutes of moderate activity five days/week or 20 minutes of vigorous activity three days/week (1). Similarly, in a research context, physical fitness-- including aerobic capacity-- is associated with morbidity, mortality and health related quality of life. Physicians and other health professionals are encouraged to screen for the level of fitness in ambulatory care settings but this is often done only by patient self-report of physical activity (1). Objective measurement is often necessary to adequately adjust for important clinical outcomes in both observational and clinical studies.

Although self-reported measures of fitness are typically adequate in population studies, they may be less useful or valid for individual clinical care. Instruments that measure physical activity through self-report often suffer from problems of recall bias, with patients frequently over-rating their activity levels (2). In addition, self-reports of physical fitness do not correlate well with more objective measures of fitness (3). An objective fitness screening test that could be used in a community setting would provide an objective measure of a patient's fitness level and may indicate a need for referral to fitness professionals. The fitness screening test could also be used in research settings where adjustment for this important issue is frequently needed.

The gold standard measure of aerobic fitness is the measurement of VO2peak (4). However, these tests are time consuming, costly, and impractical for use in community settings. Other fitness tests have been developed for screening or as field tests for population studies that are easier to implement, have a low risk of adverse events, and require less time and equipment. These approaches include timed walk tests, timed distance tests, and step tests. One of the most frequently used tests is the Six Minute Walk Test. It has been used extensively in populations with chronic disease (e.g. lung disease and heart failure) and in the elderly (5). Guidelines for this test recommend a 30-meter hallway that is seldom traveled (6). Another commonly used test, the 400-meter walk, has been used in high functioning older adult populations and recommends a 20-meter hallway and generally takes from 3 to 6 minutes (7). The space needed to provide an obstacle-free performance area and the time for completion limit the utility of both of these walking tests. In addition, heart rates achieved with walking tests regularly fall below the best range for predicting aerobic capacity in most instances (8).

Step tests are an alternative to level walking tests that require less space and modest amounts of equipment. There are a plethora of different step tests described that use different step heights and stepping rates making it difficult to compare results. Step tests may be single stage, as short as 3 minutes, or conducted in multiple stages lasting up to 20 minutes. For use both as a component of epidemiological research or in clinical settings, shorter tests--if accurate--are preferred (9).

The common, validated and reliable step tests may not be that useful in clinical

and broad population settings because of their difficulty. The measurement properties of these tests have primarily been examined in young or physically active populations. All tests use step heights significantly higher than a standard stair (7.5 inches). The three most common prescribe a stepping cadence for step climbing that approaches the average cadence for level walking. The Chester step test uses an 11.8 inch step at 60-140 steps/minute (10), the Harvard step test uses a 13.8 inch step, 96 steps/minute (11), and Queens College Step Test (QCST) uses a 16.25 inch step, 88-96 steps/min (12). All of these tests may be too difficult for middle-aged and older adults and may even be problematic for younger populations. As an example, D'Alonzo reported attrition of 25% in a study of female students with an average age of 27 due to the difficulty of the QCST (13). This study reports on the modification, use, and attrition rates of the QCST for use in a clinical setting with middle-aged adults. It is part of an epidemiological study assessing measurement equivalence in subjective survey responses and comparison with a battery of physical performance measures.

METHODS

Sample

The subjects for this report were 1,234 people aged 45-64 who were recruited as part of a study of racial and ethnic differences in subjective and objective measures of health. Aim 1 of our main study was to investigate measurement equivalence in the Short Form-36v2 (SF-36v2) across both racial and ethnic groups focusing on English and Spanish-speaking Hispanics. The methodological approach to this AIM involved sophisticated statistical analysis using structural equation modelling and confirmatory factor analysis to assess psychometric property differences in responses to the SF-36v2 across four racial/ethnic groups, paying particular attention to differences due to language translation and other cultural issues. Aim 2 of the study was to investigate how these subjective responses to the SF-36v2 compared with performance-based measures of physical functioning across these same groups. It was this segment of our study that led to our discoveries regarding the difficulties with one of our performance-based measures. By design, the study included oversampling of African Americans and people of Hispanic background. The project's target was to recruit equal numbers of non-Hispanic whites and African Americans and two similar sized Hispanic groups who did and did not speak Spanish as their primary language.

Participants were recruited from two academic medical practices and three community clinics in Cleveland, Ohio and Chicago, Illinois. Both locations had bilingual (English/Spanish) research assistants on their teams. Subjects were recruited at the clinical sites as they waited for an appointment and through public advertising (posters and newsletter announcements) at the clinical sites. A research assistant would describe the project to potential participants, determine eligibility, and inquire about interest in participation. After obtaining informed consent, each subject was scheduled to either participate that day or a future date. At the scheduled time subjects were escorted to a dedicated testing area at each site where they completed a set of questionnaires and a battery of performance-based measures of physical functioning. Institutional review boards in Chicago and Cleveland approved the study protocol.

Patients were excluded if they were non-ambulatory, had a body mass index >35 kg/m^2 or did not speak English or Spanish. Patients were screened for ability to perform

safely the physical assessment component of the study by first checking vital signs. Patients were excluded if their resting heart rate was <56 or >90 (14), resting respiratory rate >17, or resting BP >160/100 (moderate hypertension). Next, potential participants were asked a series of questions from the Physical Activity Readiness Questionnaire developed by the Canadian Society of Exercise Physiology to determine ability to exercise safely (15). Patients who answered "yes" to any of the seven screening questions were asked to get written approval from their physician prior to enrolling in the study.

Performance-based measures

Our timed step test was embedded in a study assessing racial/ethnic differences in subjective measures of health and performance based measures of physical functioning. Twelve performance-based measures of physical functioning (PBM) were chosen to measure physical function in specific domains including aerobic capacity or fitness, upper and lower extremity strength, muscle endurance, balance, dexterity, and flexibility, domains that could be directly compared to those assessed in the SF-36v2 physical functioning subscale. In choosing these tests, we tried to balance acceptability to subjects, ease of performance, evidence of test reliability, low rates of subject inability to complete the test, time required to perform the test, and the ability of the test to provide a substantial variance in the non-disabled population. A short description of 11 tests from this battery of 12 performance-based measures follows below after which we present a detailed description of the modified step-test we employed.

Upper extremity strength was assessed with a grip strength and push-pull dynamometers (16, 17). Lower extremity strength was assessed using a modified leg extension dynamometer test described by Sherrington & Lord (18). Upper extremity endurance was measured by subjects repeatedly lifting a weight (4 pounds for women, 8 for men) with the preferred arm and counting the number of lifts completed in 30 seconds (19). Lower extremity endurance was assessed with Jones, Rikli & Beam's 30-second chair-stand test (20). Timed fast gait speed was evaluated on a 10-meter hard floor surface walking area (8). Stair climbing was assessed with a timed, three repetition three stair ascent/descent (21). Upper extremity flexibility was assessed by the back scratch test (22). Lower extremity flexibility was assessed using the Flex-Tester (Novel Products, Inc., Rockton, IL) sit and reach flexibility test box (23). Balance was assessed by the one leg standing balance test (24). Dexterity was assessed with a modified Jebsen Hand Function test (25).

Aerobic Capacity: Step Test Procedure

The step test used in this study was an adaptation of the test described by McCardle (12). The original test (validated among college students), asks participants to step up onto a 16.25-inch step at a predetermined stepping rate. The required cadence rate is 96 for men and 88 for women (24 and 22 cycles per minute for men and women, respectively). A metronome was used to facilitate the subject keeping the target cadence. At the end of 3 minutes the heart rate is counted for 15 seconds, multiplied times 4 and recorded as beats per minute. However, during our pilot testing, we noted that only the fittest of middle-aged adults were able to complete the test. Because of this we first decided to lower the step height to 10 inches which is the step height used by Siconolfi in work in a middle-aged population (26). This modification significantly increased the rates of completion for women but we still found the cadence too challenging for men. Subsequently we settled on a cadence of 92 for both men and women. This step rate is between the rates identified

by the McCardle test i.e. Queen's College Step Test and were also used in the Dundee Step test (27) and Astrand protocols (28).

The research assistant demonstrated the stepping procedure in time with the metronome for each subject. To enhance patient safety we monitored heart rate throughout the test using a pulse oximeter and stopped the testing if the subject exceeded 80% of their predicted maximum heart rate, calculated using the formula [208 - (0.7 x age)] x 0.8 (29). The subjects were asked to report any nausea, chest pain, or dizziness to the research assistant who was instructed to stop the test immediately. Heart rate was recorded at one-minute intervals. People who were unable to complete the test because of general pain or fatigue were categorized as voluntarily stopping. Subjects who were instructed to stop for any reason (nausea, chest pain, dizziness, heart rate) were categorized as maximum heart rate reached. The research assistant also recorded the time point within the three-minute test at which subject participation stopped.

Health-related Variables:

We also included two health-related variables that may have influenced performance on the step test. We calculated subjects' body mass index using self-reported height and weight, and also performed a count of the number of medications that they were currently using as noted in their electronic medical record. Body mass index was categorized as normal (<25.0), overweight (25.0-29.99) and obese (30.0-35.0). The number of medications used was categorized as none, 1-4, and 5 or more.

Statistical Analysis:

The characteristics of the study population and associated univariate and bivariate statistics were generated using SAS 9.1 (SAS Institute Inc., Cary, NC) and Stata 11 (Stata Corp., College Station, TX). Comparisons across categories of subject participation/performance in the step test (i.e., refused, voluntarily stopped, maximum heart rate reached, completed) and the single continuous variable of age used analysis of variance and Bonferroni corrected tests for significance. Comparisons of step test categories and nominal or categorical variables (i.e., sex, race/ethnicity, body mass index category, number of medications category) used conventional tabulation procedures and Pearson's χ^2 test of significance.

RESULTS

Figure 1 presents a detailed flow chart of the 1,234 enrolled subjects and their attrition by time interval and reason for stopping. Using our modified step height, 160 subjects (13%) refused to even try the test, and only 350 subjects (28%) were able to complete it. A total of 280 subjects (23%) voluntarily stopped during the three-minute test and 444 (36%) were stopped by the research assistants because they had reached 80% of their predicted maximal heart rate. One person required treatment for orthostatic hypotension that developed during the testing procedure.

The demographic and health-related characteristics of the 1,234 people enrolled in the study are described in Table 1, second panel, labeled "pooled". In subsequent panels we test for differences across the step test participation categories with demographic and health related variables. To summarize the pooled findings the sample had slightly more females than males, had more overweight than either normal or obese persons and over 50% of subjects were taking no medications. By design Hispanics were over-sampled contributing to the rather large

proportion of both English and Spanish speaking Hispanics.

The cross-tabulation results show that those who completed the test were slightly younger. Higher refusals and dropout rates were found for females, Spanish speaking Hispanics and those in the obese category of body mass index (BMI). Blacks had the largest percentage of voluntary stops. Males were almost twice as likely as females to complete the test. Those not taking any medications were much more likely to complete the test.

As an expansion of Figure 1, Figures 2 and 3 are graphical presentations of participation and dropout by sex and race/ethnicity/language category across stopping time and reason for stopping (30).

Figure 1. Subject entry and retention diagram

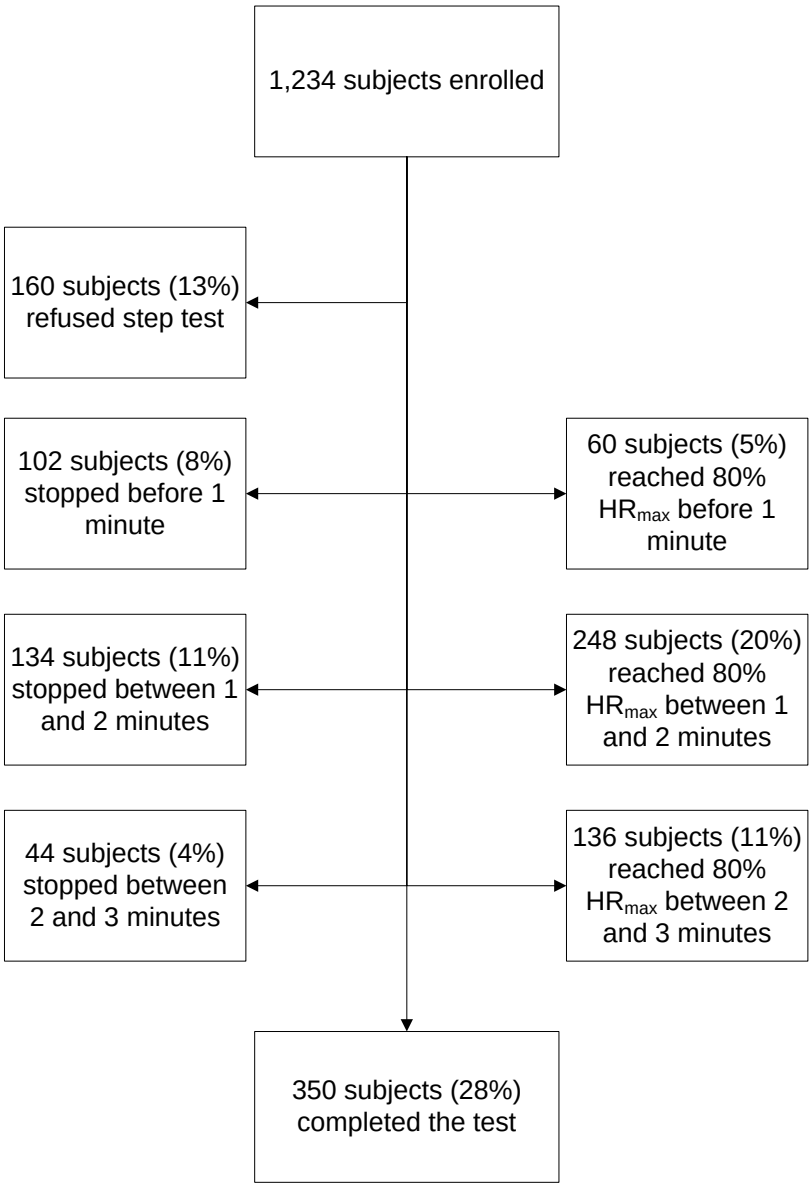


Table 1. Study Population Characteristics by Step Test Participation Category

Characteristic	Pooled (N=1234)	Refused (N=160)	Voluntary Stop (N=280)	Max Heart Rate Stop (N=444)	Completed (N=350)
Age, mean years (SD)	52.9 (5.5)	53.3 (5.4)	53.7 (5.8)	53.0 (5.6)	51.9***† (5.2)
Sex, %‡					
Male (n=548)	44.4	41.2	27.1	40.1	65.1
Female (n=686)	55.6	58.8	72.9	59.9	34.9
Race/Eth/Lang, %‡					
White (n=336)	27.2	19.4	19.3	30.2	33.4
Black (n=368)	29.8	20.6	45.4	25.9	26.6
Hispanic-Eng (n=206)	16.7	13.7	9.6	20.0	19.4
Hispanic-Span (n=324)	26.3	46.3	25.7	23.9	20.6
BMI, %‡					
Normal (n=364)	29.5	22.5	23.2	27.0	40.9
Overweight (n=484)	39.2	29.4	34.3	42.8	43.1
Obese (n=386)	31.3	48.1	42.5	30.2	16.0
Medications, %‡					
None (n=620)	50.3	38.7	50.0	48.6	57.7
One-Four (n=357)	28.9	29.4	22.5	33.2	28.6
Five or more (n=257)	20.8	31.9	27.5	18.2	13.7

Notes: All categories within columns sum to 100%. SD = Standard Deviation

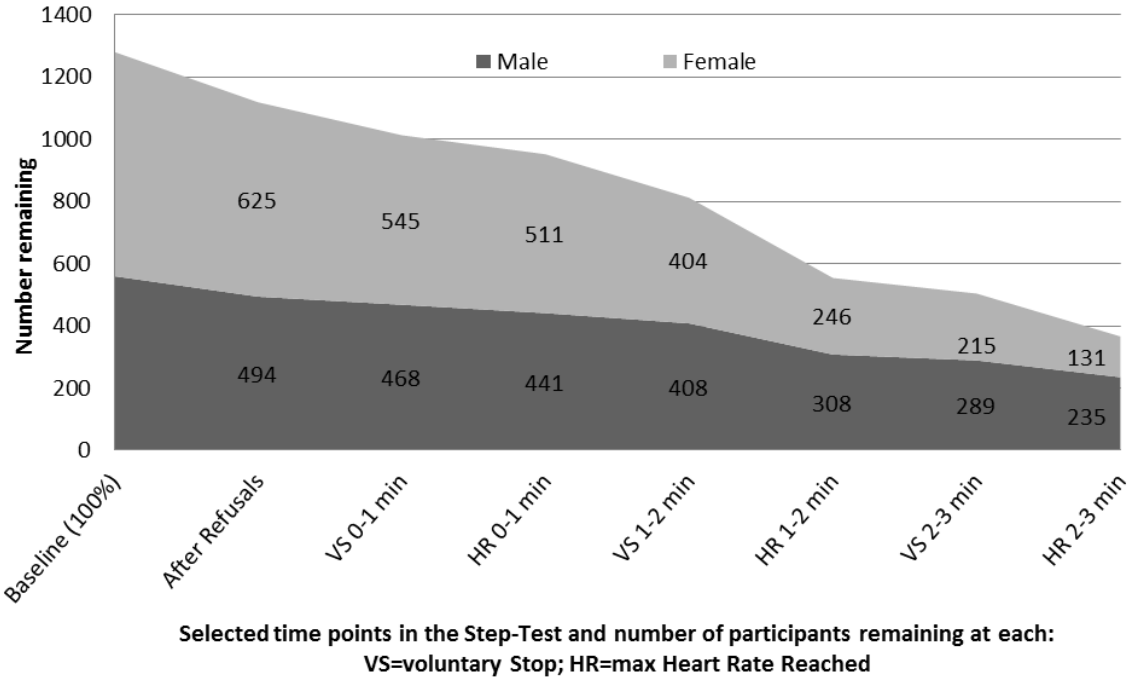
***= $p < 0.001$

† Statistical significance is compared to voluntary stop using analysis of variance.

‡ Statistical significance is for trends across all categories of the variable using Pearson χ^2 and p values were < 0.001 for sex, race/ethnicity/language, BMI and medications.

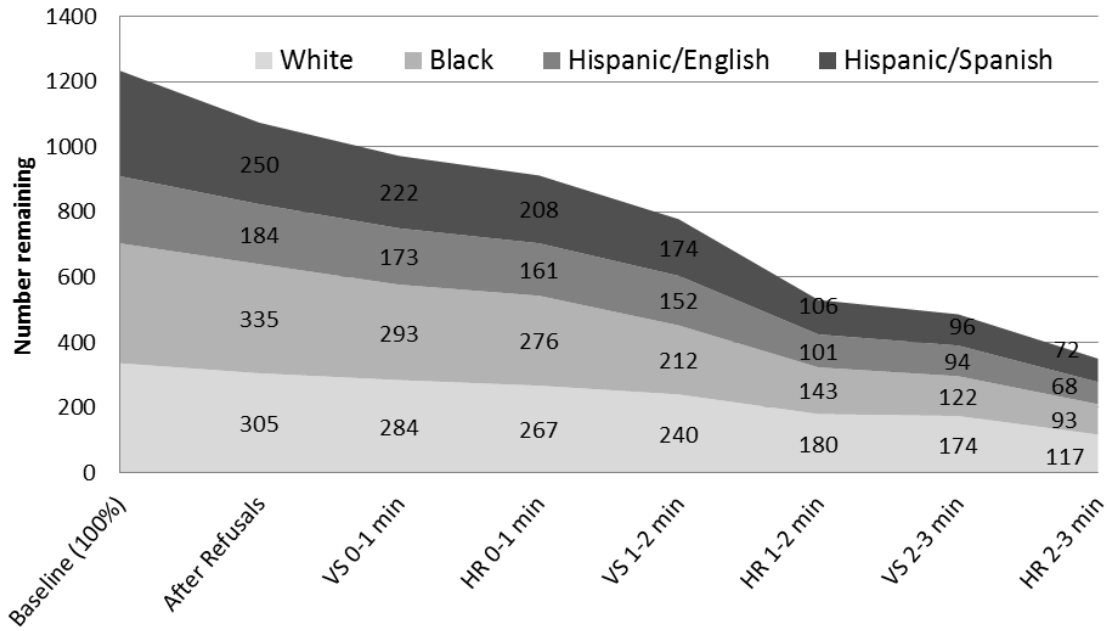
For BMI, <25.0=normal, 25.0-29.99=overweight, 30.0-35.0=obese.

Figure 2: Sex differences in step-test participation and completion.



Selected time points in the Step-Test and number of participants remaining at each:
VS=voluntary Stop; HR=max Heart Rate Reached

Figure 3: Racial/ethnic differences in step-test participation and completion.



Selected time points in the Step-Test and number of participants remaining at each:
VS=voluntary Stop; HR=max Heart Rate reached

DISCUSSION

To be useful across both clinical and epidemiologic research settings, an ideal aerobic assessment test should: 1) be safe; 2) be relatively short in duration; 3) not require expensive equipment; 4) be validated for specific disease populations; 5) be able to be completed by people with a diversity of fitness levels, impairments and ages , and 6) have good psychometric properties including high reliability and validity with respect to the gold standard assessment of aerobic capacity, the measurement of peak oxygen consumption.

In this study of a large, diverse, middle-aged clinical population, the step test we employed to assess aerobic capacity could not be completed by most subjects. Moreover, 36% of subjects had to be stopped because their heart rate exceeded our *a priori* safety parameters. This is despite an alteration of the standard test to make it easier through the use of a shorter step height and decreased cadence for men. Clearly--as implemented in this study--the step test was not useful for estimating aerobic capacity among the vast majority of this middle-aged clinical population, even after several attempts at modification, adherence to nearly all of the ideal characteristics for an aerobic assessment test as mentioned above, the exclusion of the most obese subjects (BMI > 35) and persons with potential cardiac symptoms.

Petrella, in his work in an aging population, has approached the issue of testing from a different perspective. He realized that the timed walking tests that were being used in this population frequently did not raise the heart rate to the level where VO2peak could be reliably measured (31, 32). To address this problem, he incorporated stepping into the protocol, but retained the self-paced aspect of the test. In his validation study, he used two 7.5 inch steps which correspond to a typical house step. One difference between Petrella's test and ours is the number of leg movements per cycle: six for the two step test as opposed to four for our one step test (32). Another difference is that subjects were instructed to step up and down 20 times at a "normal" pace and then 5 minutes later to do it again at a "fast" pace whereas our subjects used a cadence of 92 in a test designed to last three minutes. In spite of these differences comparing his study to ours provides some useful information. When asked to perform the task at a "normal" speed, Petrella's subjects (average age = 72) used a cadence of 57 and 59 steps per minute for men and women, respectively. When asked to step "fast" the cadence increased to 65 and subjects were able to keep this pace for approximately two minutes. While the task in our step test and Petrella's is different because of the difference in steps (2 versus 1), even when subjects were asked to step "fast" , the cadence chosen was much slower than that prescribed for the QCST test (96 for men and 88 for women) and much slower than the 92 cadence we used in our study. We hypothesize that the higher cadence prescribed in our test (compared to Petrella's self-paced test) may be a major reason for the dropout rate in our study. The self-paced alternative would presumably allow many more subjects to be able to complete the test as they could slow their cadence as they tired.

While our test is clearly too difficult for many subjects, prior studies and our work performed in this study indicates that a self-paced test with a step height between 8 and 10 inches might allow for high completion rates, heart rates above 122 beats per minute to be reached safely, monitoring of fitness and an administration time period that is workable in population research or clinical settings Such a tool would be useful in clinical settings to monitor the response to

exercise programs and could be as routine as taking blood pressure and equally as important to maintain health (33).

Our subgroup analysis by age, sex, racial/ethnic grouping, BMI and number of medications indicates that there may be differences in dropout rates related to race, ethnicity, and sex. Further work controlling for other covariates would be required to better assess whether these differences are real or are related to other selection differences in the subjects enrolled. For example, there may be large difference in disease burdens between these groups that might explain differential dropout rates. Similarly, physical differences between men and women in body structure, stamina and other factors might explain sex differentials. Controlling for factors such as we mentioned above would most likely eliminate the vast majority of differences seen in our study, but future research is needed to address these and other limitations in our work.

Limitations

One limitation in our study is the exclusion of a large number of people, primarily related to obesity. In this study we chose to limit the sample to those with a BMI less than 35. It is unlikely that including persons with higher BMIs would change the conclusion concerning the lack of utility of the test we used for assessing aerobic capacity.

Second, we used self-reported weight and height. Prior research suggests we underestimated obesity prevalence which may have contributed to the lack of significance of BMI categories to the explained variance in dropout rates.

Third, our subjects were between the ages of 45 and 64 and therefore our results should not be generalized to other age categories or populations.

Fourth, our study population is a clinical one. Therefore, any generalizations to community dwelling individuals should be made with caution.

CONCLUSIONS

This study performed among older adults ages 45 to 64 indicates that existing fixed cadence step tests that are not graded are particularly problematic for assessing aerobic capacity/fitness in a clinical population because of high dropout rates during testing. Further exploration of self-paced step tests should focus on the ability of subjects to complete the tests and the validity of those tests to predict VO2peak. Self-paced step tests typically used in elderly populations should also be evaluated for utility and validity among younger populations.

FUNDING

This work was supported by the National Institute on Aging at the National Institutes of Health, Grant No. R01-AG022459.

ACKNOWLEDGEMENTS

The authors want to acklowledge writing/editing and technical help from Sandra Andrukat and Misty Harris, MPH.

REFERENCES

1. Haskell WL, Lee I-M, Pate RR, Powell KE, Blair SN, Franklin BA, Macera CA, Heath GW, Thompson PD, Bauman A. Physical activity and public health. Circulation. 2007; 116:1081-93.

2. Warms C. Physical activity measurement in persons with chronic and disabling conditions. Fam Community Health. 2005; 29:785-885.

3. Heimmel J, Patel S, Cody R, Bachmann G. Evaluation of physical fitness in an ambulatory setting. Am J Obstet Gynecol. 2007; 196:522.e1-524.e4.

4. American College of Sports Medicine (ACSM). ACSM's Guidelines for Exercise Testing and Prescription (9th ed). Philadelphia: Lippincott, Williams & Wilkins. 2013.

5. Enright PL. The six-minute walk test. Respir Care. 2003 Aug;48(8):783-5.

6. American Thoracic Society (ATS). ATS Statement: Guidelines for the six-minute walk test. Am J Respir Crit Care Med. 2002;166:111-117.

7. Simonsick EM. Fan E. Fleg JL. Estimating cardiorespiratory fitness in well-functioning older adults: treadmill validation of the long distance corridor walk. J Am Geriatr Soc . 2006; 54(1):127-32.

8. Astrand P-O, Rodahl K, Dahl HA, Stromme SB, . Textbook of Work Physiology, 4th edition. New York: Mcgraw-Hill, 2003.

9. VanSwearingen JM and Brach JS..Making geriatric assessment work: selecting useful measures. Phys Ther 2001;81:1233-1252.

10. Buckley JP, Sim J, Eston RG, Hession R, Fox R. Reliability and validity of measures taken during the Chester step test to predict aerobic power and to prescribe aerobic exercise. Br J Sports Med. 2004;38:197-205.

11. Chen S, Wang J, Lee W, HOu C, Chen C, Laio Y, Lin C, Kuo C. Validity of the 3 min step test in moderate altitude: environmental temperature as a confounder. Appl Physiol Nutr Metab. 2006;31:726-730.

12. McArdle WD, Katch FI, Pechar GS, Jacobson L, Ruck S. Reliability and interrelationships between maximal oxygen intake, physical work capacity and step-test scores in college women. Med Sci Sports. 1972;4(4):182–186.

13. D'Alonzo KT, Stevenson JS, Davis SE. Outcomes of a program to enhance exercise self-efficacy and improve fitness in Black and Hispanic college-age women. Res Nurs Health. 2004 Oct;27(5):357-69.

14. Fox K, Borer JS, Camm AJ, et al. Resting heart rate in cardiovascular disease. J Am Coll Cardiol. 2007; 50: 823-830.

15. Canadian Society for Exercise Physiology. Patient Activity Readiness Questionnaire (PAR-Q). Accessed June 24, 2013 at: www.csep.ca/cmfiles/publications/parq/par-q.pdf

16. Smidt N, van der Windt DA, Assendelft WJ, Mourits AJ, Deville WL, deWinter AF, Bouter LM. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. Arch Phys Med Rehabil. 2002; 83:1145-1150.

17. National Institute for Occupational Safety and Health (NIOSH). Pre-employment strength testing. Washington, DC: US Department of Health and Human Services. 1977.

18. Sherrington C, Lord SR. Reliability of simple portable tests of physical performance in older people after hip fracture. Clin Rehabil. 2005 Aug;19(5):496-504.

19. Malmberg JJ, Miilunpalo SI, Vuori IM, Pasanen M, Oja P, and Haapanen-Niemi N. A health-related fitness and functional performance test battery for middle-aged and older adults: feasibility and health-related content validity. Arch Phys Med Rehabil. 2002;83:666-677.

20. Jones CJ, Rikli RE, Beam WC. A 30-second chair-stand test as a measure of lower body strength in community-residing older adults. Res Q Exerc Sport. 1999; 70:113-119.

21. Lechner DE, Page JJ, Sheffield G. Predictive validity of a functional capacity evaluation:The physical work performance evaluation. Work. 2008;31(1):21-25.

22. Rikli RE, Jones CJ. Development and validation of a functional fitness test for community-residing older adults. J Aging Phy Act. 1999;7:129-161.

23. Golding LA. YMCA Fitness Testing and Assessment Manual. 4th ed. Champaign, IL: Human Kinetics. 2000.

24. Suni JH, Oja P, et al. Health-related fitness test battery for adults: aspects of reliability. Arch Phys Med Rehabil. 1996;77:399-405.

25. Bovend'Eerdt TJH, Dawes H, Johansen-Berg H, Wade DT. Evaluation of the modified Jebsen test of hand function anf the University of Maryland Arm questionnaire for stroke. Clin Rehabil. 18:195-202 (2004)

26. Siconolfi SF, Garber CE, Lasater TM, Carleton RA. A simple, valid step test for estimating maximal oxygen uptake in epidemiologic studies. Am J Epidemiol. 1985;121:382-390.

27. Lim P, Shiels P, Anderson J, MacDonald T. Dundee step test: a simple method of measuring the blood pressure response to exercise. J Hum Hypertens. 1999 Aug; 13(8):521-6.

28. Tammelin T, Nayha S, Rintamaki H. Cardiorespiratory fitness of males and females of northern Finland birth cohort of 1966 at age 31. Int J Sports Med. 2004; 25(7):547-552.

29. Tanaka, H., Monahan, K. D., Seals, D. R. Age-Predicted Maximal Heart Rate Revisited. J Am Coll Cardiol 2001; 37:53-6.

30. Tufte, ER. The Visual Display of Quantitative information. Cheshire, CT: Graphics Press. 1986.

31. Petrella RJ, Wight D. An office-based instrument for exercise counseling and prescription in primary care. Arch Fam Med. 2000; 9:339-344.

32. Petrella RJ, Koval JJ, Cuningham DA, Paterson DH. Can primary care doctors prescribe exercise to improve fitness? The Step Test Exercise Prescription (STEP) project. J Prev Med. 2003; 24: 316-322.

33. Petrella RJ, Koval JJ, Cuningham DA, Paterson DH. A self-paced step test to predict aerobic fitness in older adults in the primary care clinic. J Am Geriatr Soc. 2001;49:632-638.