Patterns of leisure-time physical activity using multivariate finite mixture modeling and cardiovascular risk factors in the Northern Manhattan Study

Ying Kuen Cheung, PhD[a,*], Gary Yu, DrPH[a], Melanie M. Wall, PhD[a], Ralph L. Sacco, MD, MS[b], Mitchell S. V. Elkind, MD, MS[c,d], and Joshua Z. Willey, MD, MS[c]

[a]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY

[b]Department of Neurology and Public Health Sciences, Miller School of Medicine, University of Miami, Miami FL

[c]Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY

[d]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY

Abstract

Purpose—Physical activity is currently commonly summarized by simple composite scores of total activity, such as total metabolic equivalent score (METS), without further information about the many specific aspects of activities. We sought to identify more comprehensive physical activity patterns, and their association with cardiovascular disease risk factors.

Methods—The Northern Manhattan Study is a multiethnic cohort of stroke-free individuals. Questionnaires were used to capture multiple dimensions of leisure-time physical activity. Participants were grouped into METS categories, and also into clusters by multivariate mixture modeling of activity frequency, duration, energy expenditure, and number of activity types. Associations between clusters and risk factors were assessed using chi-squared tests.

Results—sing data available in 3293 participants, we identified six model-based clusters that were differentiated by frequency and diversity of activities, rather than activity duration. High activity clusters had lower prevalence of the risk factors compared to those with lower activity; associations with obesity and hypertension remained significant after adjusting for METS (p = .027, .043). METS and risk factors were not significantly associated after adjusting for the clusters.

Conclusions—Data-driven clustering method is a principled, generalizable approach to depict physical activity and form subgroups associated with cardiovascular risk factors independently of METS.

Keywords

Cluster analysis; Exercise; Hypertension; Metabolic equivalent; Obesity; Questionnaires

Introduction

Leisure-time physical activity is an important component of primary prevention for cardiovascular disease and stroke across all age groups [1–4]. As physical activity and diet are both complex behaviors, the measurements to assess these behaviors are necessarily multidimensional. However, the existing literature and recommendations focus on measuring physical activity by some pre-defined summary statistic such as total metabolic equivalent score, energy expenditure or duration of exercise. The American Heart Association guidelines for primary prevention of cardiovascular disease recommend 150 minutes of moderate intensity or 75 minutes of heavy intensity activity per week [5]. While these recommendations consider both time and intensity of the activities, they leave several unanswered questions as to how physical activities should be carried out to achieve optimal health outcomes in other aspects including frequency and number of different types of activities based on data. For example, though 150 minutes per week are recommended for moderate intensity activity, it is not clear if there is a difference between 150 minutes in one session, versus five sessions of 30 minutes each. In contrast, analytical methodology for dietary assessment has evolved from using a single summary index such as total energy intake to applying cluster analysis to raw data from food frequency questionnaires, so that dietary patterns are identified in a data-driven manner [6, 7].

The multivariate finite mixture modeling (MFMM) analysis is a model-based, data-driven clustering method. The underlying assumption of cluster analysis is that the entire cohort consists of a mixture of subgroups, where the number of subgroups is not known. By explicitly making model assumptions on the data, the MFMM aims to produce subgroups of subjects with similar exercise patterns arising from the same statistical distribution, and to divide the cohort into an optimal number of subgroups based on some specific criteria [8]. For model-based methods, the principle is general enough to encompass different types of data; however, much attention has been paid to clustering algorithms for data arising from multivariate normal distributions [9, 10] or dichotomous Bernoulli distributions where it is often called latent class analysis [11]. Data collected in physical activity questionnaires are of mixed type including skewed continuous and count measures. Our specific objectives were to execute the MFMM cluster analysis for multivariate data of mixed type, explore whether it would identify meaningful patterns and detailed description of leisure-time physical activity, and assess the association between the physical activity patterns and cardiovascular disease risk factors. Our eventual goal was to generate information that could allow providers to give more specific counseling to patients beyond a single index of total weekly activity.

Material and methods

Study population

The Northern Manhattan Study (NOMAS) is a population-based cohort study designed to evaluate the effects of medical, socio-economic, and other risk factors on the incidence of vascular disease in a stroke-free multiethnic community-based cohort. Methods of participant recruitment, evaluation and follow-up have been previously reported [12]. In-person evaluations were performed at Columbia University Medical Center or at home for those who could not come in person 6% were performed at home). The study was approved by the institutional review boards at Columbia University Medical Center and the University of Miami. All participants gave informed consent to participate in the study.

The NOMAS cohort consists of a total of 3,298 participants recruited between 1993 and 2001 with mean age 69 years at baseline, 63% women, and 52% Hispanic. Baseline physical activity questionnaire data were available in 3,293 subjects.

Data collection

Physical activity was measured by an interviewer-administered questionnaire adapted from the National Health Interview Survey of the National Center for Health Statistics [13]. The questionnaire consisted of 14 pre-specified types of leisure-time physical activity. The 14 items were walking, jogging or running, hiking, gardening or yard work, aerobics or aerobic dancing, other dancing, calisthenics or general exercise, golf, tennis, bowling, bicycle riding, swimming or water exercise, horseback riding, handball, racquetball, or squash. In addition, two open fields were allowed for additional activities that were not listed. We kept the two other activities separate to account for the diversity of the different activities performed. For each activity, each of the following self-report variables was asked: the participation in the activity type, the frequency that each activity was conducted within the past two-week period, and the duration of conduct of the activity at each session. The questionnaire has been previously reported as reliable for individuals reporting moderate physical activity and validated in this population, demonstrating a crude concordance rate of 0.69 when proxies of the participants were asked [13]. The same measure also correlated with body mass index, activities of daily living scores, and quality of well-being activity scores [14].

Questionnaires were linked with compendia of physical activity to allow calculation of metabolic equivalents (MET) [kcal/kg-hour] for the intensity of activity as well as energy expenditure in kilocalories [15]. The metabolic equivalent score of an activity by a participant was calculated as the product of MET and duration of the activity; the total metabolic equivalent score (METS) of a participant was calculated by summing the products across all activities. In previous analyses, total physical activity was classified based on quartiles of METS into three subgroups: METS<1, METS between 1 and 14, and METS>14 [16]. In this study, in addition to METS, we also measured total physical activity by the total frequency of leisure-time physical activities in the past two weeks regardless of types, the average duration per session (calculated as the total duration of activities divided by the total frequency), the number of types of activities conducted, and the total energy expenditure (calculated as the product of METS and the participant's body weight).

Data regarding baseline status and risk factors were collected through interviews of participants. Race-ethnicity was determined by self-identification. Standardized questions were asked regarding the following conditions: hypertension, diabetes, cigarette smoking, and cardiac condition. Standard techniques were used to measure height, weight, and blood pressure. Diabetes mellitus was defined as fasting blood glucose 126mg/dL or higher, the patient's self report of diabetes mellitus, or insulin or hypoglycemic agent use. Hypertension was defined as systolic blood pressure 140 mmHg or diastolic blood 90 mmHg based on the average of 2 blood pressure measurements, physician diagnosis, or patient self-report of a history hypertension or antihypertensive use. Obesity was defined as BMI 30. High waist circumference was defined as >40cm for men, and >35cm for women.

Statistical analysis

We performed cluster analysis using multivariate finite mixture modeling (MFMM) in participants who reported any physical activity. Specifically, the inputs to the algorithm include four physical activity measures: the frequency of any physical activity, the duration of any activity, the number of activity types, and the total energy expenditure due to any activity. The analysis was conducted assuming normality of the logarithm of frequency, duration, and energy expenditure, while the number of activity types as count data was modeled using Poisson distribution. The analysis assumes a common variance-covariance matrix of the lognormal variables across clusters; and conditional independence of the Poisson variable given cluster membership. Individuals who reported no physical activity were grouped into a separate cluster. The choice of the number of clusters relied on the Bayesian information criterion (BIC) [17] with a maximum of 6 clusters. Estimates of the model parameters were obtained using maximum likelihood in Mplus, version 6.11.

Physical activity patterns, demographic and baseline status variables, and cardiovascular risk factors of each identified cluster were described using means and standard deviations for continuous variables, and proportions for categorical variables. While the inputs to the clustering algorithm were the four physical activity measures given above, we also looked at other physical activity measures within each cluster, including walking patterns (e.g., the frequency of walking per week).

Associations with clusters were assessed using chi-squared tests for categorical variables, and ANOVA for continuous variable (e.g. age). When and only when a global chi-squared test found significant association between a baseline variable and clusters, chi-squared tests would be used to compare clusters to the no-activity group.

Associations between risk factors and clusters were also assessed using chi-squared tests after excluding the no-activity group. Specifically the risk factors considered were diabetes, hypertension, cardiac disease, obesity, and high waist circumference at baseline. We assessed the association of risk factors and clusters with adjustments for the METS categories using logistic regression: the presence of risk factors was used as the dependent variable with METS category and cluster as independent variables. Inference was based on likelihood ratio tests. The association tests and logistic regression were performed in R, version 3.0.1.

Results

Summary of physical activities in the NOMAS cohort

Table 1 summarizes the leisure-time physical activity patterns in the entire NOMAS cohort in terms of average duration, frequency, and energy expenditure per week. A total of 1,971 participants reported at least some physical activities within two weeks of interviews, whereas 1,322 reported no activity. A total of 2,742 activities were reported among the 1,971 participants, indicating that some participants were engaged in more than one activity. Among these participants, the mean duration of physical activities was 44 minutes per session with a mean of 5.9 sessions per week and the mean energy expenditure was 1,300 kcal per week. Walking was the principal activity reported.

Physical activity patterns by multivariate finite mixture modeling

The MFMM found a five-cluster solution based on BIC among the 1971 subjects with at least some activities. Table 2 reports a summary of physical activity measures in each cluster based on any activity and walking only. Successively more active clusters were characterized by higher frequency of activities and larger number of activity types, rather than longer duration per session. The standard deviation of activity frequency in each cluster was much reduced from that in the entire cohort (SD=3.9) indicating the clusters consist of homogeneous subgroups in terms of activity frequency. Highly active participants (cluster VI) reported on average 2 sessions per day with a mean 35 minutes per session, whereas the active daily cluster (cluster V) was active approximately once a day with a mean of 45 minutes each. In addition, highly active participants (cluster VI) tended to have increased number of activity types. Participants in clusters V and VI have very similar walking patterns in terms of duration (walked 48 minutes on average vs. 45 minutes), frequency (walked 6 days a week on average vs. 7 days), and energy expenditure (spent 1.3 kcal per week on walking vs. 1.4 kcal per week), but are distinguished by the number of activity. Cluster II was characterized by rare activity (about once per 2 weeks) but with long per-session duration (mean=60 minutes); however, this cluster contained few subjects (n=74) and had low activity level overall with only 3% meeting the American Heart Association guidelines of total activity. In comparison, 71% of cluster V and 97% cluster VI met the goals with respective weekly averages of 312 and 507 minutes of exercise per week.

There was a statistically significant association between the MFMM clusters and the three METS categories (P < 0.001); see Table 3. The MFMM refined grouping of subjects within each METS category. In the intermediate METS category (between 1 and 14), subjects were reclassified mainly into three moderate-sized clusters (III, IV, V); and subjects in the high METS category (>14) were mainly placed in clusters V and VI.

Physical activity patterns, demographic and baseline status variables, and risk factors

Table 4 summarizes the demographic and baseline status variables by clusters. There was no significant difference in the mean age among the MFMM clusters. The distributions of sex, race-ethnicity, education, smoking status, alcohol consumption, and social support were highly significantly different among the clusters (P 0.001). In particular, the highly active group (cluster VI) was significantly different from the no-activity group (cluster I) on these

variables: Cluster VI was characterized by higher proportions of men (46%), whites (38%), completing at least high school (68%), former smokers (51%), moderate alcohol consumption (52%), and social support of more than 3 friends (93%) when compared to the other clusters. Not surprisingly, except for alcohol consumption, the rare-activity-group (Cluster II) was similar to the no-activity group.

Table 5 summarizes the distributions of cardiovascular risk factors by clusters. There were significant associations between the MFMM clusters and diabetes, hypertension, obesity, and high waist circumference. The association with hypertension (P=0.043) and obesity (P=0.027) remained significant after adjusting for the METS. In particular, Cluster VI was characterized by a lower proportion of hypertension; and those in cluster V and cluster VI were less likely to be obese when compared to the other groups.

Discussion

Using the MFMM analysis based on multiple physical activity measures (instead of a single score to summarize total activity such as METS), we found that the highly active group (Cluster VI, characterized by higher activity frequency and more activity types), was associated with lower prevalence of hypertension and obesity even after adjusting for the conventional METS category. We performed a similar association analysis between METS categories and risk factors, and found that the METS was not associated with risk factors after adjusting for the MFMM clusters. Thus, our approach produced subgroups that were associated with risk factors independently of METS, while the conventional METS categories would provide no additional information once conditioned on the MFMM clusters. And the improvement in association was not only due to refinement of the METS categories. For example, the subgroup in cluster IV with METS>14 (n = 60) had an average METS of 29 and 30% obesity, whereas the subgroup in cluster V with METS between 1 and 14 (n = 509) had an average METS of 10 and 24% obesity. That is, the association of the clusters and obesity may not be achieved by simply using a finer gradation based on METS alone.

In our cohort, the frequency of activities and the number of activity types were the two main descriptive factors of physical activity patterns, as opposed to per-session duration. This information could be useful in prescribing recommendations to patients beyond total weekly activity. Interestingly, our approach identified a subgroup with rare activity but long duration per session (Cluster II) that has similar total weekly activity to Cluster III. If we had used total activity as the only descriptive variable, we would not have identified these as two distinct subgroups. In theory, comparing Clusters II and III may provide specific answers to whether it is equivalent or not to exercise frequently with a moderate amount or to exercise rarely with long duration. Unfortunately, as these two clusters were characterized by low activity level and the cluster sizes were relatively small (n=74, 196), our analysis did not provide information to address this question using this particular data set. However, the identification of these two clusters demonstrates the potential use of MFMM analysis to explore physical activity patterns in a data-driven manner. Likewise, our approach identified two subgroups with very similar walking patterns (Clusters V and VI) but were clearly separated based on the number of types of activities. Therefore, although walking was the

major physical activity in this cohort, our approach was able to utilize information from additional activities.

Though traditionally physical activity is thought to have no upper limit for health benefits, recent literature suggests potential adverse health outcomes with higher and more extreme levels. One study found that extremely vigorous weight-bearing exercise as compared to its more moderate counterpart resulted in lower bone density with the possibility for osteoporosis in individuals after the age of 50 [18], while others reported an increased risk of injury, such as cardiac fibrosis, associated with strenuous excessive exercise [19, 20]. Our analysis, on the other hand, did not find an upper limit and suggested the active subgroups (Clusters V, VI) were associated with favorable risk profiles. Our results are in line with previous investigators who examined the effects of physical activity on reduced mortality and extended life expectancy [21]. Our analysis sheds new insight on this debate in that health benefits due to intense physical activity could be achieved through higher frequency and variety of exercise, rather than long and irregular single sessions. One possible explanation is that frequent and regular exercise sessions with moderate duration is associated with self-discipline of the exercisers, which may in turn be associated with other healthy behaviors.

The MFMM analysis has some important strength. First, it is data-driven and can accommodate multiple clustering variables thus addressing the multi-dimensionality of physical activity data. It avoids the use of a single summary index for total activity. Second, the principle can be applied to other high dimensional data types such as activity meters, and be generalized to other populations so as to account for local variability in life-space, neighborhood characteristics, socio-demographic factors, and could allow for the inclusion of baseline co-morbidities into the information used to define each cluster [22, 23]. Third, when compared to two standard heuristic clustering approaches, namely K-mean clustering [24] and hierarchical agglomerative method, the MFMM analysis is a principled method based on statistical models. K-mean clustering is based on iterative partitioning by which data points are classified from one group to another until there is no further improvement based on the sum of squares criterion; a drawback of this method is that it assumes that the number K of clusters is known. In hierarchical agglomerative clustering, two groups with minimum distance are merged at each step of the iteration with each data point being its own group at the start of the algorithm; the output of this iteration is a hierarchy of similarity (called dendrogram) between data points and there is no clear way to determine the number of clusters. In contrast, because the MFMM analysis is model-based, likelihood-based criteria can be applied to determine the optimal number of subgroups. In particular, we chose BIC as our optimization criterion because simulation study suggests it performs best among other criteria [25]. A popular alternative is the Akaike information criterion, which penalizes model complexity less than BIC and is thus less parsimonious.

Our study has also some limitations. First, our analyses with socio-demographic factors and cardiovascular disease risk factors were cross-sectional and as such we could not draw on conclusion about the directionality of association. Future analyses will include association with longitudinal clinical outcomes [26]. Second, there is incomplete information in our cohort regarding non-leisure time activity, such as occupational and commuting activity.

While it is possible that participants who are highly active as part of their employment would not perform leisure-time physical activity, several studies have reported an independent protective effect on cardiovascular disease from leisure-time physical activity, independent of other forms of physical activity [23, 27–32]. However, the majority of the cohort was not working at the time of enrollment. Third, due to the elderly population, the range of physical activity may be limited and may not provide the information to assess any harmful effects of over-exercising as in the previous reports [18–20]. Fourth, the MFMM analysis in theory can take in as many clustering variables as the raw data has, and can be completely data-driven. In practice, we often need to do some dimension reduction based on subject knowledge to avoid computational and convergence difficulties. Our analysis used four physical activity measures (frequency, average duration, activity types, and energy expenditure) derived from the 15 questionnaire items. While in principle we could apply MFMM analysis to frequency, average duration, and energy expenditure of each of the 15 items, the use of many data dimensions would cause computation to be prohibitive. In addition, sparse data with many zeros would likely cause convergence problems in fitting the maximum likelihood estimate. Alternative clustering approaches such as hierarchical agglomerative methods would have the difficulty with time efficiency, as it requires memory usage proportional to the square of the sample size. Having said this, our study has illustrated the usefulness of MFMM analysis as an exploratory tool for complex behavior patterns. Our analysis represents an important improvement upon the conventional over-simplistic approach of using a single index.

Conclusion

The MFMM analysis yields description of physical activity patterns in a data-driven manner. The patterns are associated with cardiovascular risk factors independently of METS in a cross-sectional analysis. Further prospectively analyses with risk factors and outcomes are warranted.

Acknowledgments

List of abbreviations

MET	Metabolic equivalents
METS	Total MET score
MFMM	Multivariate finite mixture modeling
NOMAS	Northern Manhattan Study

References

1. Thompson PD, Buchner D, Pina IL, Balady GJ, Williams MA, Marcus BH, et al. Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease: a statement from the Council on Clinical Cardiology (Subcommittee on Exercise, Rehabilitation, and

Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity). Circulation. 2003; 107:3109–16. [PubMed: 12821592]

2. Nelson ME, Rejeski WJ, Blair SN, Duncan PW, Judge JO, King AC, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. Circulation. 2007; 116:1094–105. [PubMed: 17671236]

3. Kiely DK, Wolf PA, Cupples LA, Beiser AS, Kannel WB. Physical activity and stroke risk: the Framingham Study. Am J Epidemiol. 1994; 140:608–20. [PubMed: 7942761]

4. Lee CD, Folsom AR, Blair SN. Physical activity and stroke risk: a meta-analysis. Stroke. 2003; 34:2475–81. [PubMed: 14500932]

5. Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, et al. Heart disease and stroke statistics–2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation. 2008; 117:e25– e146. [PubMed: 18086926]

6. Hu FB, Rimm E, Smith-Warner SA, et al. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. Am J Clin Nutr Feb. 1999; 69(2):243–9.

7. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol Feb. 2002; 13(1):3–9.

8. McLachlan, G.; Peel, D. Finite Mixture Models. Wiley; 2000.

9. Banfield JD, Raftery AE. Model-based Gaussian and non-Gaussian clustering. Biometrics. 1993; 49:803–21.

10. Fraley, C.; Raftery, AE.; Murphy, B.; Scrucca, L. Technical Report No. 597. Department of Statistics, University of Washington; 2012. mclust version 4 for R. mclust version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation.

11. Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. Stat Med. 1986; 5:21–7. [PubMed: 3961312]

12. Sacco RL, Anand K, Lee HS, Boden-Albala B, Stabler S, Allen R, et al. Homocysteine and the risk of ischemic stroke in a triethnic cohort: the Northern Manhattan Study. Stroke. 2004; 35:2263–2269. [PubMed: 15345803]

13. Moss AJ, Parsons VL. Current estimates from the National Health Interview Survey: United States, 1985. Vital Health Stat. 1987; 10:i–iv. 1–182.

14. Sacco RL, Gan R, Boden-Albala B, Lin IF, Kargman DE, Hauser WA, et al. Leisure-time physical activity and ischemic stroke risk: the Northern Manhattan Stroke Study. Stroke. 1998; 29:380–7. [PubMed: 9472878]

15. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Medicine and Science in Sports Exercise. 2000; (32):S498–S504.

16. Willey JZ, Moon YP, Paik MC, Yoshita M, DeCarli C, Sacco RL, et al. Lower prevalence of silent brain infarcts in the physically active: the Northern Manhattan Study. Neurology. 2011; 76(24): 2112–8. [PubMed: 21653889]

17. Fraley C, Raftery AE. How Many Clusters? Which Clustering Method? Answers Via Model Based Cluster Analysis. The Computer Journal. 1998; 41(8):578–589.

18. Michel BA, Bloch DA, Fries JF. Weight-Bearing Exercise, Overexercise and Lumbar Bone Density Over Age 50 Years. Arch Intern Med. 1989; 149:2325–9. [PubMed: 2802897]

19. Möhlenkamp S, Lehmann N, Breuchmann F, Bröcker-Preuss M, Nassenstein K, Halle M, et al. Running: the risk of coronary events. Prevalence and prognostic relevance of coronary atherosclerosis in marathon runners. European Heart Journal. 2008; 29:1903–10. [PubMed: 18426850]

20. Benito B, Gay-Jordi G, Serrano-Mollar A. Cardiac Arrhythmogenic Remodeling in a Rat Model of Long-Term Intensive Exercise Training. Circulation. 2011; 123:13–22. [PubMed: 21173356]

21. Wen CP, Wai JP, Tsai MK. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. Lancet. 2011; 278(9798):1244–53. [PubMed: 21846575]

22. Norton MC, Dew J, Smith H, Fauth E, Piercy KW, Breitner JC, et al. Lifestyle behavior pattern predicts incident dementia and Alzheimer's Disease. The Cache County Study. J Am Geriatr Soc. 2012; 60(3):405–12. [PubMed: 22316091]

23. Holtermann A, Marott JL, Gyntelberg F, Søgaard K, Suadicani P, Mortensen OS, et al. Does the benefit on survival from leisure time physical activity depend on physical activity at work? A prospective cohort study. PLoS One. 2013; 8(1):e54548. [PubMed: 23349926]

24. MacQueen, J. Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability; 1967; Berkely, Calif: University of California Press; p. 281-97.

25. Nylund KL, Asparouhov T, Muthen BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Structural Equation Modeling. 2007; 14:535–569.

26. Willey JZ, Moon YP, Paik MC, Boden-Albala B, Sacco RL, Elkind MSV. Physical activity and risk of ischemic stroke in the Northern Manhattan Study. Neurology. 2009; 73:1774–9. [PubMed: 19933979]

27. Hu G, Eriksson J, Barengo NC, Lakka TA, Valle TT, Nissinen A, et al. Occupational, commuting, and leisure-time physical activity in relation to total and cardiovascular mortality among Finnish subjects with type 2 diabetes. Circulation. 2004; 110(6):666–73. [PubMed: 15277321]

28. Hu G, Tuomilehto J, Borodulin K, Jousilahti P. The joint associations of occupational, commuting, and leisure-time physical activity, and the Framingham risk score on the 10-year risk of coronary heart disease. Eur Heart J. 2007; 28(4):492–8. [PubMed: 17242011]

29. Hu G, Jousilahti P, Borodulin K, Barengo NC, Lakka TA, Nissinen A, et al. Occupational, commuting and leisure-time physical activity in relation to coronary heart disease among middle-aged Finnish men and women. Atherosclerosis. 2007; 194(2):490–7. [PubMed: 16979645]

30. Hu G, Sarti C, Jousilahti P, Silventoinen K, Barengo NC, Tuomilehto J. Leisure time, occupational, and commuting physical activity and the risk of stroke. Stroke. 2005; 36(9):1994–9. [PubMed: 16081862]

31. Clays E, De Bacquer D, Janssens H, De Clercq B, Casini A, Braeckman L, et al. The association between leisure time physical activity and coronary heart disease among men with different physical work demands: a prospective cohort study. Eur J Epidemiol. 2013; 28(3):241–7. [PubMed: 23329153]

32. Sisson SB, Camhi SM, Church TS, Martin CK, Tudor-Locke C, Bouchard C, et al. Leisure time sedentary behavior, occupational/domestic physical activity, and metabolic syndrome in U.S. men and women. Metab Syndr Relat Disord. 2009; 7(6):529–36. [PubMed: 19900152]

Table 1

Summary of the fifteen physical activity items in the Northern Manhattan Study

Activity	N	Ave minutes per session Mean (SD)	Frequency per week Mean (SD)	Total kcal per week (10^3) Mean (SD)
1. Walking	1636	46 (38)	4.9 (2.5)	1.1 (1.2)
2. Jogging or Running	51	36 (22)	3.2 (1.9)	1.0 (0.8)
3. Hiking	11	164 (130)	0.7 (0.3)	0.7 (0.6)
4. Gardening or Yard Work	32	74 (90)	2.4 (2.3)	1.0 (1.7)
5. Aerobics or Aerobic Dancing	100	36 (27)	3.2 (2.2)	0.7 (1.0)
6. Other Dancing	67	87 (91)	1.6 (1.8)	0.7 (1.0)
7. Calisthenics or General Exercise	476	23 (17)	4.4 (2.6)	0.5 (0.6)
8. Golf	14	210 (137)	1.3 (0.9)	1.4 (1.5)
9. Tennis	6	88 (62)	0.9 (0.7)	0.6 (0.4)
10. Bowling	7	74 (46)	2.1 (2.4)	0.30 (0.1)
11. Bicycle Riding	95	42 (50)	3.6 (2.6)	0.8 (0.9)
12. Swimming or Water Exercises	63	52 (35)	2.1 (1.3)	0.8 (0.7)
13. Horseback riding	0	--	--	--
14. Handball, Racquetball, or Squash	2	60 (0)	0.8 (0.4)	0.6 (0.4)
15a. Other Activity 1	171	59 (119)	2.6 (2.0)	0.9 (2.2)
15b. Other Activity 2	11	66 (65)	2.1 (1.8)	1.3 (2.6)
Overall	1971	44 (44)	5.9 (3.9)	1.3 (1.5)

Table 2

Five physical activity patterns in the Northern Manhattan Study by multivariate finite mixture modeling

	I: No activity (n = 1322)	II: Rare activity (n=74) [a]	III: Active weekly (n=196)	IV: Active every other day (n=472)	V: Active daily (n=1035)	VI: Highly active (n=194)
Any activity						
Ave minutes per session	0 (0)	60 (62)	37 (28)	46 (58)	45 (39)	35 (22)
Frequency (per week)	0 (0)	0.5 (0)	1.2 (0.3)	2.9 (0.7)	7.0 (1.2)	14 (3)
Total kcal (10^3) per week	0 (0)	0.16 (0.2)	0.25 (0.2)	0.73 (1.2)	1.6 (1.4)	2.8 (2.0)
Number of activity types	0 (0)	1.0 (0)	1.0 (0.2)	1.2 (0.5)	1.4 (0.6)	2.4 (0.9)
Meeting AHA guidelines[a]	0 (0)	3%	3%	26%	71%	97%
Walking only						
Ave minutes per session	0 (0)	50 (55)	39 (28)	45 (35)	48 (40)	45 (32)
Frequency (per week)	0 (0)	0.5 (0)	1.2 (0.3)	2.7 (0.9)	6.0 (1.7)	7.0 (2.3)
Total kcal (10^3) per week	0 (0)	0.07 (0.1)	0.16 (0.2)	0.44 (0.5)	1.3 (1.3)	1.4 (1.2)

[a] 150 minutes/week moderate exercise

Except for "Meeting AHA guidelines", each entry is mean (sd)

Table 3

Physical activity patterns using multivariate finite mixture modeling and metabolic equivalent score (METS) categories in the Northern Manhattan Study

	I: No activity (n=1322)	II: Rare activity (n=74)	III: Active weekly (n=196)	IV: Active every other day (n=472)	V: Active daily (n=1035)	VI: Highly active (n=194)
METS category, n						
< 1	1322	7	6	11	0	0
1–14	0	66	188	401	509	11
>14	0	1	2	60	526	183
Total MET-hrs/week, mean (sd)	0 (0)	2 (3)	3 (3)	10 (13)	22 (19)	39 (24)

Table 4

Association of physical activity patterns using multivariate finite mixture modeling and baseline demographics in the Northern Manhattan Study

	[a] Cluster						P
	I	II	III	IV	V	VI	
Age							
Mean	69	68	69	69	69	70	.34
SD	10	10	10	10	10	10	
Sex							
Females %	67	57	67	64	58	54	<.001
Males %	33	43	33	36	42	46	
Race-Ethnicity							
Whites %	16	11	18	21	26	38	<.001
Blacks %	21	28	24	23	31	27	
Hispanics %	63	61	58	56	43	35	
Education							
No high school %	62	62	64	53	47	32	<.001
At least high school %	38	38	36	47	53	68	
Smoking Status							
Never smoker %	47	49	55	51	46	38	<.001
Former smoker %	35	39	29	35	36	51	
Current smoker %	18	12	16	14	18	11	
Moderate Alcohol Consumption							
No %	73	58	65	66	65	48	<.001
Yes %	27	42	35	34	35	52	
[b]**Social Support**							
No %	18	16	14	15	13	7	.001
Yes %	82	84	86	85	87	93	

[a] Cluster I: No activity (n = 1322); II: Rare activity (n = 74); III: Active weekly (n = 196); IV: Active every other day (n = 472); V: Active daily (n = 1035); VI: Highly active (n = 194).

[b] Defined as having three or more friends

Table 5

Association of physical activity patterns using multivariate finite mixture modeling and cardiovascular risk factors in the Northern Manhattan Study

	[a] Cluster						[b] P	[c] P_{adj}
	I	II	III	IV	V	VI		
Diabetes								
No %	76	77	78	76	81	84	.012	.68
Yes %	24	23	22	24	19	16		
[d] Odds ratio	1.00	0.94	0.91	1.00	0.74*	0.58*		
Hypertension								
No %	24	26	22	27	28	37	.004	.043
Yes %	76	74	78	73	72	63		
[d] Odds ratio	1.00	0.93	1.15	0.87	0.85	0.55*		
Cardiac Disease								
No %	78	77	80	79	79	79	.94	.98
Yes %	22	23	20	21	21	21		
[d] Odds ratio	1.00	1.04	0.87	0.93	0.92	0.91		
Obese								
No %	69	74	66	70	77	80	<.001	.027
Yes %	31	26	34	30	23	20		
[d] Odds ratio	1.00	0.81	1.21	0.98	0.71*	0.59*		
High Waist Circumference								
No %	53	61	53	55	61	68	<.001	.42
Yes %	47	39	47	45	39	32		
[d] Odds ratio	1.00	0.77	1.00	0.90	0.74*	0.56*		

[a] Cluster I: No activity (n = 1322); II: Rare activity (n = 74); III: Active weekly (n = 196); IV: Active every other day (n = 472); V: Active daily (n = 1035); VI: Highly active (n = 194).

[b] P values of global chi-squared tests for association between clusters and risk factors

[c] P values of likelihood ratio tests for association after adjusting for METS category

[d] Odds ratio with risk factors using Cluster I as reference. An "*" indicates odds ratio is statistical significantly different from 1