Self-selection contributes significantly to the lower adiposity of faster, longer-distanced, male and female walkers.

Paul T. Williams[1]

Abbreviations: BMI: body mass index; km kilometer; m/s meters per second

[1] Life Sciences Division, Lawrence Berkeley Laboratory, Donner Laboratory, Berkeley, CA. 94720 ptwilliams@lbl.gov

Keywords: Methodology, bias, self-selection, cross-sectional studies, observational studies, statistical association, exercise, walking, body mass index, adiposity.

Abstract

Although cross-sectional studies show active individuals are leaner than their sedentary counterparts, it remains to be determined to what extent this is due to initially leaner men and women choosing to exercise longer and more intensely (self-selection bias). In this report walking volume (weekly distance) and intensity (speed) were compared to current BMI ($BMI_{current}$) and BMI at the start of walking ($BMI_{starting}$) in 20,353 women and 5,174 men who had walked regularly for exercise for 7.2 and 10.6 years, respectively. The relationships of $BMI_{current}$ and $BMI_{starting}$ with distance and intensity were nonlinear (convex). On average, $BMI_{starting}$ explained >70% of the association between $BMI_{current}$ and intensity, and 40% and 17% of the association between $BMI_{current}$ and distance in women and men, respectively. Although the declines in $BMI_{current}$ with distance and intensity were greater among fatter than leaner individuals, the portions attributable to $BMI_{starting}$ remained relatively constant regardless of fatness. Thus self-selection bias accounts for most of the decline in BMI with walking intensity and smaller albeit significant proportions of the decline with distance. This demonstration of self-selection is germane to other cross-sectional comparisons in epidemiological research, given self-selection is unlikely to be limited to weight or peculiar to physical activity.

Thirty minutes of moderate-intensity physical activity on most days of the week has been recommended by the Surgeon General and other government and nongovernment organizations for significant health benefits (1-3), and one-hour per day for the prevention of unhealthy weight gain (4,5). Walking is specifically advocated by these guidelines (1-4) and national survey data show it is the most commonly practiced exercise in the United States (5).

Cross-sectional studies (6-11) show physically active individuals are leaner than their sedentary counterparts. By comparing cross-sectionally total energy expenditure between ideal-weight and overweight men, the Institute of Medicine derived their exercise recommendations for preventing unhealthy weight gain (4). Although there is also experimental evidence showing that weight loss can be induced by exercise (12-16), these are limited by both small sample size and durations of usually less than a year. In contrast, published cross-sectional studies often involve thousands of subjects whose weights reflect the long-term consequences of physical activity and concomitant health behaviors (8-11).

Cross-sectional associations may not prove causality but often it is presumed logically and from other evidence. Specifically, the cross-sectional association between physical activity and body weight is usually attributed to the effect of exercise because activity requires energy expenditure and because animals and people lose weight when exercised (17,18), however this does not preclude the additional effect of heavier individuals choosing to be less active (19-21). Although this caveat is usually acknowledged, there is in fact a paucity of documentation of this particular self-selection bias.

In this paper, we derive quantitative estimates of the proportion of the inverse association between walking distance and body weight that can be attributed to initially leaner men and women choosing to walk faster and further. Specifically, data collected as part of the National Walkers’ Health Study included indices of current total and regional adiposity and recollections of these indices prior to walking regularly for exercise [8]. These are compared to current walking intensity (speed) and volume (km/wk) to calculate the percentages of the declines in current weight that are due to pre-exercise weight, i.e. self-selection. The principles established herein are germane to other cross-sectional comparisons, given self-selection is unlikely to be limited to weight or peculiar to physical activity.

Materials and Methods

A two-page questionnaire was mailed to walkers identified through a walking magazine subscriber list [8]. Walking volume was taken as the participant's usual weekly walking distance for the year the survey was completed. Walking intensity was the participant's reply to the survey question "During your usual walk, how many minutes does it take for you to walk one mile?". Body mass index (BMI) was calculated from self-reported height and weight. We also asked the participants to "Please provide, to the best of your ability, your body circumference in inches" without further instruction. The relationships between circumference and walking distance will be weakened by different perception of where waist, hip and chest circumferences lie. However, unless the perceived location varies systematically in relation to distance or speed, this subjectivity is unlikely to produce spurious relationships with walking quantity or intensity. Bra-cup sizes were coded on a five-point scale from 1 (A cup), 2 (B cup), 3 (C cup), 4 (D cup), and 5 (E cup or larger). The survey also solicited information on demographics (age, race, education), age when began walking at least 12 miles per week, frequency of walks, longest walk, weight history (weight and body circumferences when started walking 12 or more miles per week, at greatest weight, and when 18 years old); diet (vegetarianism and the current weekly intakes of alcohol, red meat, fish, fruit; vitamin C, vitamin E, and aspirin), current and past cigarette use, prior history of heart attacks and cancer, and medications for blood pressure, thyroid, cholesterol and diabetes. The study protocol was reviewed by the University of California Berkeley Committee for the Protection of Human Subjects, and all subjects provided a signed statement of informed consent.

The reproducibility of the self-reported weight and body dimensions were assessed in 3,209 subjects (2,603 females, 606 males) who submitted two survey questionnaires (mean (SD)) 1.0 (1.6) years apart. The repeated surveys showed self-reports of current weights and dimensions were correlated r=0.96 for body weight, 0.89 for waist circumference, r=0.79 for hip circumference, r=0.88 for chest circumference, and r=0.88 for bra cup size. The correlations for recollections of starting values (i.e., weights and dimensions when they began walking 12 or more miles per week) were 0.91 for body weight, 0.86 for starting waist circumference, r=0.87 for starting hip circumference, r=0.83 for starting chest circumference, and r=0.85 for starting bra cup size. Regression analysis was used to test whether of changes in the recollections of starting values were related to changes in current values adjusted for sex. Specifically, we were concerned that recollections of starting weight would be simply reflections of current weight less a few pounds. Changes in current weight or body dimensions between the first and second survey did not appear to significantly affect recollections of starting values for body weight (P=0.62), or circumferences of the waist (P=0.67), hip (P=0.19), or chest (P=0.46) in the 3,206 subjects, whereas changes in women's current bra cup size significantly influenced their recollections starting bra cup size (p <0.0001).

Statistical analyses

Age-adjusted walking distance and speed, BMI, circumference measures, and bra cup sizes were produced as the residuals from sex-specific quadratic least-squares regression. Walking distance and speed were compared to current BMI ($BMI_{current}$) and BMI when participants first began walking 12 or more miles per week ($BMI_{starting}$), and to current and starting waist, hip and chest circumferences, and bra cup sizes. We previously demonstrated that the relationships of women's BMI and body circumferences to walking distance were nonlinear (convex)[8], and Figure 1 shows that the relationships of women's BMI to walking speed are convex as well. We therefore used standard polynomial regression with adiposity measures as the dependent variable and walking distance (km/wk) and walking distance squared (km/wk^2) as the independent variables. From these equations, the change in adiposity corresponding to a one km/wk or m/s increment in walking distance or speed (i.e., from X to X+1 m/s) was calculated as the tangent to the regression curve, which is given by first derivative of the polynomial regression equation evaluated at velocity X (Figure 1). The percentages attributable to self-selection were computed from the ratio of the slopes for $BMI_{starting}$ and $BMI_{current}$ evaluated at selected distances and speeds, and the weighted average of these ratios based on the distribution of distance and velocity in the sample, where we specified a maximum of 100 if the ratio exceeded one and a minimum of 0 percent when the ratio was less than zero.

In addition to being convex, we have also previously demonstrated that the regression slopes relating female adiposity to walking distance depend upon the sample percentile of the BMI or body dimension [8]. This is also true for walking velocity (Figure 1). Specifically Figure 1 shows that the 90^{th} percentile of BMI declines more rapidly with exercise velocity than lower BMI percentiles, a phenomenon also proven for exercise distance in female walkers [8], and in both male and female runners using similar methodology [9-11]. In this report, nearest neighbors were used to determine the percentiles of the dependent variable corresponding to each X_i, i=1..N. Specifically, the bivariate observations (X_i,Y_i) were ordered from smallest to largest X to yield the ordered set of observations $(X_{[i]},Y_i)$. We then sorted values of the dependent variable from the one hundred nearest walking distances to $X_{[i]}$ from smallest to largest. These sorted values were used as the $1^{st}(Y_{i[1]})$, $2nd(Y_{i[2]})$, $3rd\ (Y_{i[3]})$, ...100th $((Y_{i[100]}))$ percentile of adiposity corresponding to the walking distance $X_{[i]}$. Quadratic polynomial regression was applied to $(X_i,\ Y_{i[k]})$ to estimate the relationship of weekly walking distance to the k^{th} percentile of BMI. The tangent slopes were calculated at selected velocities (e.g., 1.2 m/s, 1.8 m/s, and 2.4 m/s, Figure 1) and plotted as functions of percentiles as done elsewhere [8-12]. We also computed the average ratio of the slopes for starting and current weight over all percentiles and across all distances or velocities using weights based on sample distribution of distances or speeds.

Results

Of the 7,364 men and 28,376 women who were nonsmokers and nonvegetarians who did not use medications for thyroid conditions or diabetes, 70% of men and 72% of women provided complete information on height and weight so that BMI could be calculated at the time the survey was completed and when they first walked 12 miles per week. Most of those excluded did not provide their weight when they first started walking 12 miles or more per week. This was either because they never achieved that level of exercise or because they could not recall their earlier weights. The former presumably accounts for most of the exclusions because: 1) those missing starting weights walked less than half the weekly distances as those providing these data [mean (SE): men 11.1 (0.2) vs 24.8 (0.2) km/wk; women 11.2 (0.1) vs. 23.9 (0.1) km/wk]; and 2) 92% of those missing starting weights provided other historical weight data. Those providing both starting and current weights were generally middle-aged or older [means (SD), men: 60.5 (12.7), women: 49.2 (12.5) years], with some college education [men: 16.0 (2.8), women: 15.0 (2.5) years], marginally overweight [BMI men: 26.9 (4.4); women: 25.2 (5.0) kg/m^2], and usually walked 1.81 (0.51) m/s if male and 1.81 (0.44) m/s if female. Self-reported body circumferences of the waist, hip and chest averaged 93.4 (10.1) cm, 101.0 (10.2) cm, and 107.4 (10.2) cm respectively in men, and 77.5 (11.5) cm, 99.2 (10.6) cm, and 93.4 (8.3) cm respectively in women. On average, the men had walked 12 or more miles per week for 10.7 (10.3) years and the women for 7.2 (7.4) years.

Self-selection and exercise volume Figure 2 presents the histogram of the mean difference in current BMI ($BMI_{current}$) between the least active women (i.e., 1st decile of weekly walking distance) and women of the 2nd through 10th decile of current walking distance. These differences increase with progressively greater distance. However, the corresponding differences in average starting BMI ($BMI_{starting}$) also increased progressively with current walking distance, accounting for approximately half of the $BMI_{current}$ differences.

Table 1 presents the quadratic regression equations relating $BMI_{current}$ and $BMI_{starting}$ to walking distance. For completeness, we also reported the regression coefficients for the difference between $BMI_{current}$ and $BMI_{starting}$. The statistical significances of the quadratic terms confirm the nonlinearity of the relationships. The table also presents the formulas for the tangent slopes, which are used to estimate the expected effect of a one km/wk change in walking distance on BMI starting at “0 km/wk” (i.e., from being completely sedentary to walking one km/wk, column 4), 20 km/wk (column 5), and 40 km/wk (column 6, see Figure 1 for example of tangents). These are presented because quadratic regression equations are themselves enigmatic, whereas the tangent slopes at any particular point is readily interpreted as the estimated change in adiposity for a one km increment in weekly distance walked. The tangent slopes will be the same for all walking distances when the relationship is linear, and will become less negative with increasing distance when the relationship is convex. The regression equations for $BMI_{current}$, $BMI_{starting}$, and □□□□□ □□□□□□□□□□□ are all strongly convex, i.e., their tangent slopes at 0 km/wk are at least four-fold greater than their slope at 40 km/wk. Below the tangent slopes are the percents of the relationships attributable to self-selection, which are calculated from the ratio of the slopes for $BMI_{starting}$ to $BMI_{current}$ {i.e.,100*(slope for $BMI_{starting}$/slope for $BMI_{current}$}. The percentage attributable to self selection is only slightly less at 20 km/wk than 0 km/wk, but diminishes by nearly half by 40 km/wk. When averaged over the distribution of women’s walking distances, the regression analysis attributes 40.4% of the slope for $BMI_{current}$ to self-selection.

Figure 3 displays the tangent slopes for the 5th through 95th percentiles of $BMI_{current}$ versus walking distance. The tangents are evaluated at 20 km/wk. The Y-axes of figure 3 are the slopes and the X-axes their percentiles. The graph shows that per km/wk walked, the decline in $BMI_{current}$ became progressively greater as the percentile of $BMI_{current}$ increased. The graph also show that per km/wk walked, the decline in $BMI_{starting}$ became progressively greater with the percentile of $BMI_{starting,.}$ The ratio of the slopes $BMI_{starting}$ to $BMI_{current}$ varies unsystematically about the value of 36%. Thus although the slope for the 95th percentile of $BMI_{current}$ vs walking distance is twelve-fold greater than the slope at the 5th percentile, remarkably the proportion attributed to self selection remains relatively constant. When averaged over all percentiles and weighted by the women's distribution of walking distances, 36.8% of the decline in $BMI_{current}$ with distance can be attributed to the decline in $BMI_{starting}$.

Women's waist, hips, and chest circumferences, and their reported bra cup sizes also declined in association with walking distance (Figures 2 and 4). Whereas substantial proportions of the declines in hip and chest circumferences are attributable to pre-exercise levels, declines in bra cup and waist circumference are not. Table 1 shows that the relationships of pre-exercise body circumferences to walking distance were all significantly nonlinear (i.e., convex). Averaged over all individuals (see methods), self-selection accounted for 19.8%, 54,4%, 46.6% and 15.4% of the decline in waist, hip, chest, and bra sup size, respectively. At greater walking distances the contribution of self-selection diminishes along with the exercise effect (cf 40 vs 20 or 0 km/wk).

Figure 5 shows that men's $BMI_{current}$ also declined with walking distance, but only small portions of the differences between the least active men and those of the 2nd through 10th deciles of walking distance could be attributed to pre-exercise BMI. Table 3 shows that the quadratic regression coefficients were only marginally significant, and that only a small proportion of the relationship of $BMI_{current}$ to walking distance can be attributed to self-selection (16.6% when averaged over all men). Figure 3 show that as in women, the decline in men's $BMI_{current}$ per km/wk became progressively greater at higher percentiles of $BMI_{current}$. Consistent with Table 3, only a small proportion of the decline in men's $BMI_{current}$ was accounted for by $BMI_{starting}$, and that the proportion was relatively constant over the percentile distribution of $BMI_{current}$. None of the decline in men's waist circumference with walking distance was attributed to self-selection.

Self-selection and moderate-exercise intensity Figures 1,2 and 4 show usual walking speed exhibits a strong inverse relationship with all five measures of current adiposity in women, and to BMI and waist circumference in men. The associations are significantly nonlinear (i.e., significance quadratic terms for the regression analyses of Tables 2 and 4, particularly among women) and convex (i.e., the tangent slopes at 2.4 m/s being one-quarter to one-third of 1.2 m/s for all variables except men's waist circumference, which is about 40% less). Figures 1 and 3 show that the declines for $BMI_{current}$ per m/s increased substantially from the lowest to the highest percentiles of the $BMI_{current}$ distribution. In woman who walk at 1.2 m/s the decline in 95^{th} percentile of $BMI_{current}$ per m/s was over twenty fold larger than the slope for the 5^{th} $BMI_{current}$ percentile in women. It was three fold larger in men.

In women, the analyses of Table 2 show $BMI_{starting}$ and pre-exercise body circumferences and bra cup sizes are also convex, and account for nearly all of the association of current hip and chest circumferences with walking speed (average 97.4 and 98%, respectively). The table also shows that self-selection accounts for the majorities of the association of walking speed with $BMI_{current}$ (average 79.9%) and bra cup size (average 66.5%). Slightly less than half of the association of walking speed and women's waist circumferences reflect pre-exercise circumferences (average 47.8%). The proportions of current adiposity attributable to self-selection appear relatively consistent across intensities, i.e., the ratios of the slopes for starting to current adiposity are similar at 0, 20, and 40 km/wk. Figure 3 shows that at 20 km/wk, all of the associations between walking speed and 25^{th} percentile of $BMI_{current}$ and above were attributable to $BMI_{starting}$.

In men, Figure 5 and Table 4 show that pre-exercise values were the major determinent of the association between walking speed and $BMI_{current}$ (average 70.7%) and Figure 3 verifies this finding for the 5^{th} through the 95^{th} percentiles of BMI current. There is good agreement between the average proportions attributable to $BMI_{starting}$ between the 5^{th} and 95^{th} percentiles of Figure 5 and the tangent slopes at 20 km/wk in Table 4. On average, only 13.1% of the slope for current waist circumference versus walking speed was attributable to starting circumference.

Discussion

Evidence for the health benefits of walking and other moderately-intense physical activities are largely observational, linking the activity dose to indicators of disease risk (blood pressure, adiposity) or to disease risk determined prospectively [6,8,22,23, 24]. These observations are supported by experimental studies showing that health benefits are at least partially caused by physical activity [25,26,27]. Many experimental studies use vigorous-intense activity as their intervention to maximize statistical power, from which it is inferred that more moderate activity will produce a proportionally smaller effects that are nevertheless clinically significant [23]. If we are to be primarily dependent upon observational data to quantify the health benefits of moderate-intensity physical activity, then we must seek quantitative estimates of the bias that may affect these associations and modify our expectations accordingly. Although most cross-sectional studies dutifully acknowledge they have not proven causality, this caveat is principally theoretical, there are being few actual studies that quantify their potential biases.

Our findings show that the inverse associations of adiposity measures with walking are due in part to leaner men and women choosing to walk longer distances, and that most of their associations with walking intensity are due to leaner men and women choosing to walk faster. When adjusted for pre-exercise weight, there remains a significant and clinically important decline in adiposity with the dose of walking in both men and women, however, most or all of the associations between walking intensity and adiposity were eliminated by adjustment for self-selection. Our results are consistent with the observations by others that body weight is a barrier to being physically active [19] and that body weight predicts inactivity in prospective epidemiological studies [20,21]. Weight differences between active and sedentary older women have been shown to trace back to their weights during young adulthood [28]. Self-selection would explain why the relationship between adiposity and physical activity is more easily documented in cross-sectional observational studies of moderate activity than in longitudinal studies [29]. Specifically, self-selection augments cross-sectional associations but not longitudinal associations of change.

Walking may be performed at varying intensities dependent upon speed. Relative to being at rest, walking raises energy expenditure two-fold (very slow, 2 mph) to four and a half fold (4.5 mph, very very brisk)[30]. Thus walking may be classified as light intensity (i.e., activities that expend less than 3 times resting metabolic rate) to moderate intensity (i.e., three to six fold increase from resting metabolic rate), but usually not vigorous intensity (over 6-fold increase)[5]. Walking pace has been reported to reduce cardiovascular disease, i.e., relative to women averaging 0.89 to 1.34 m/s, cardiovascular disease risk decreased 24% in women who walked between 1.34 and 1.79 m/s, and by 43% in women whose pace exceeded 1.79 m/s [22]. Hard intensity walking appears to improve high-density lipoprotein cholesterol and cardiorespiratory fitness whereas moderate-intensity walking may not [31].

The relationships of walking intensity to adiposity measures have not been previously described for this sample. Our analyses suggest that this relationship is nonlinear, exhibiting a degree of convexity comparable to that previously described for BMI vs. walking distance [8]. The quadratic regression coefficients were significant for BMI and body dimensions in women, and for men's BMI. The relationship between men's waist circumference and their walking intensity became more nonlinear with the removal of starting BMI. Figures 1 and 3 show that the regression slope between $BMI_{current}$ and walking intensity increased dramatically from the leanest to the fattest individuals, i.e., the slopes for the 95^{th} percentile of $BMI_{current}$ were over twenty-fold larger than the slopes at the 5^{th} percentile in women, and over three-fold larger in men.

Adjustment for pre-exercise weights did not negate our recently-published findings of the nonlinear relationships between women's walking distance and weight [8]. Compared to the slope at 0 km/wk, without correcting for starting weight the women's slope at 40 km/wk was 83% less for BMI, 86% less for waist circumference, 75% less for circumferences of the hip and chest, and 65% less for bra cup size, whereas with correction the slopes at 40 km/wk were 69%, 72%, 44%, 41% and 49% respectively, and the quadratic term remained significant for BMI, waist circumference, and bra cup size.

The regression slopes for $BMI_{current}$ versus walking distance and intensity were substantially greater women than men. Specifically, the women's regression slope for BMI vs distance was 2.5-fold larger at 0 km/wk and two-fold larger at 20 km/wk than the men's slopes, whereas the men's and women's slopes were comparable at 40 km/wk. The slopes for BMI vs intensity were two to 2.3-fold greater in women than men. Figure 4 shows that these sex-differences persisted particularly above the 25^{th} BMI percentile (note that the scale of the women's vertical axis are 2.5-fold greater for walking distance and five-fold greater for walking intensity). However, a substantial portion of this difference can be attributed self-selection, i.e., subtracting $BMI_{starting}$ eliminated 50-70% of the male-female difference in slope for distance and about 80% of that for speed. This difference may reflect the age difference between the sample of men and women in addition to sex-effects.

Although most prospective epidemiological studies of physical activity adjust for BMI [30], this may be inadequate for two reasons. First, classical adjustment for covariates requires that the same relationship applies at all percentiles of BMI, whereas Figures 1 and 3 show there are three to twenty-fold difference from lowest to highest BMI percentiles. Thus, classical methods will over-adjust the lower percentiles of BMI and under-adjust the higher percentiles. Second, classical statistical adjustment also assumes that the covariate is determined without measurement error, and will under-adjust the data if measured imprecisely because the coefficient for the covariate will be biased toward zero [31]. The analyses of Tables 1-4 do not use $BMI_{starting}$ as a covariate but rather subtract $BMI_{starting}$ directly from $BMI_{current}$. Errors in recalling $BMI_{starting}$ (i.e., measurement error) contribute to the residual errors but they do not bias the estimate of the regression slope, and although this increases the standard errors for the regression slope, but its affect on hypothesis testing is inconsequential given our sample size.

The credibility of our findings is based in part on their derivation from large sample of men and women who have walked for exercise for many years. A subset of participants who provided repeated questionnaires affirms the reproducibility of the adiposity measures at the time that the survey was completed and the recollections of adiposity when they first began walking 12 or more miles per week. Their data also showed that recollections of pre-exercise weights and body dimensions were not biased by contemporary values (see methods). Unlike other exercises, all individuals engage in some walking since early childhood except for the physically impaired, and it was necessary to select a threshold for the commencement of walking for exercise. In retrospect, it may have been preferable to use lower threshold of activity than 12 mi/wk, which was chosen to be the same as our parallel study of runners [9-11,33], but which resulted in the exclusion of many walkers who never achieved this weekly distance. Twelve miles per week is, however, substantially less than the volume of exercise thought to be required to maintain healthy weight [4,5].

In this report, we have demonstrated the principle that self-selection may distort estimates of the health benefits of moderate-intensity physical activity, particularly those based on intensity. We believe that this principle is pertinent to the limitations of other epidemiological analyses. There remains strong compelling arguments for most Americans to increase their physical activity [1-4], however, the impact on disease risk may be significantly less than currently projected.

Grants and acknowledgements: Supported in part by grant HL-45652 and DK066738 from the National Heart Lung and Blood Institute, and was conducted at the Lawrence Berkeley Laboratory (Department of Energy DE-AC03-76SF00098 to the University of California).

1. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, Buchner D, Ettinger W, Heath GW, King AC, Kriska A, Leon AS, Marcus BH, Morris J, Paffenbarger RS, Patrick K, Pollock ML, Rippe JM, Sallis J, Wilmore JH. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. JAMA. 1995 ;273:402-7.

2. Fletcher GF, Balady G, Blair SN, Blumenthal J, Caspersen C, Chaitman B, Epstein S, Sivarajan Froelicher ES, Froelicher VF, Pina IL, et al. Statement on exercise: benefits and recommendations for physical activity programs for all Americans. A statement for health professionals by the Committee on Exercise and Cardiac Rehabilitation of the Council on Clinical Cardiology, American Heart Association. Circulation, 1996:94:857-862.

3. Physical activity and cardiovascular health. NIH Consensus Development Panel on Physical Activity and Cardiovascular Health. JAMA. 1996:276:241-246.

4. Institute of Medicine. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids (Macronutrients). The National Academies Press. Washington DC. 2002 936 pages.

5. Simpson ME, Serdula M, Galuska DA, et al. Walking trends among U.S. adults. Am. J. Prev. Med. 2003;25:95–100.

6. Chan CB, Spangler E, Valcour J, Tudor-Locke C. Cross-sectional relationship of pedometer-determined ambulatory activity to indicators of health. Obes Res. 2003;11:1563-70.

7. Ryan as, Nicklas BJ, Elahi D. A cross-sectional study on body composition and energy expenditure in women athletes during aging. Am. J. Physiol. 1996;271:E916–E921.

8. Williams PT. Nonlinear relationships between weekly walking distance and adiposity in 27,596 women. Med Sci Sports Exerc. 2005;37:1893-901.

9. Williams PT, Satariano WA. Relationships of age and weekly running distance to BMI and circumferences in 41,582 physically active women. Obes Res. 2005;13:1370-80.

10. Williams PT, Pate RR. Cross-sectional relationships of exercise and age to adiposity in 60,617 male runners. Med Sci Sports Exerc. 2005;37:1329-37.

11. Williams PT. Vigorous exercise and the population distribution of body weight. Int J Obes Relat Metab Disord. 2004;28:120-8.

12. Miller WC, Koceja DM, Hamilton EJ. A meta-analysis of the past 25 years of weight loss research using diet, exercise or diet plus exercise intervention. Int J Obes Relat Metab Disord. 1997;21:941-947.

13. Garrow JS, Summerbell CD. Meta-analysis: effect of exercise, with or without dieting, on the body composition of overweight subjects. Eur J Clin Nutr 1995;49:1-10.

14. Ross R, Dagnone D, Jones PJH, Smith H, Paddags A, Hudson R, Janssen I. Reduction in Obesity and Related Comorbid Conditions after Diet-Induced Weight Loss or Exercise-Induced Weight Loss in Men. A Randomized, Controlled Trial. Ann Intern Med. 2000;133:92-103.

15. Wood PD, Stefanick ML, Williams PT, Haskell WL.The effects on plasma lipoproteins of a prudent weight-reducing diet, with or without exercise, in overweight men and women. N Engl J Med. 1991; 325:461-466.

16. Jakicic JM, Marcus BH, Gallagher KI, Napolitano M, Lang W. Effect of exercise duration and intensity on weight loss in overweight, sedentary women: a randomized trial. JAMA. 2003;290:1323-1330.

17. Hill JO, Wyatt HR. Role of physical activity in preventing and treating obesity. J Appl Physiol. 2005;99:765-70.

18. Jakicic JM, Otto AD. Physical activity considerations for the treatment and prevention of obesity. Am J Clin Nutr. 2005 Jul;82(1 Suppl):226S-229S.

19. Ball K, Crawford D, Owen N. Too fat to exercise? Obesity as a barrier to physical activity. Aust N Z J Public Health 2000; 24: 331–333.

20. Bak H, Petersen L, Sorensen TIA. Physical activity in relation to development and maintenance of obesity in men with and without juvenile onset obesity. International Journal of Obesity 2004;28: 99–104.

21. Petersen L, P Schnohr P, Sørensen TIA Longitudinal study of the long-term relation between physical activity and obesity in adults. International Journal of Obesity 2004 28, 105–112.

22. Manson JE, Greenland P, LaCroix AZ, Stefanick ML, Mouton CP, Oberman A, Perri MG, Sheps DS, Pettinger MB, Siscovick DS. Walking compared with vigorous exercise for the prevention of cardiovascular events in women. N Engl J Med. 2002;347:716-25.

23. Erlichman J, Kerbey AL, James WP. Physical activity and its impact on health outcomes. Paper 2: Prevention of unhealthy weight gain and obesity by physical activity: an analysis of the evidence. Obes Rev. 2002; 3:273-287.

24. Ching PLYH, Willett WC, Rimm EB, Colditz GA, Gortmaker SL, Stampfer MJ. Activity level and risk of overweight in male health professionals. Am J Public Health 1996; 86: 25-30.

25. Asikainen TM, Miilunpalo S, Oja P, Rinne M, Pasanen M, Uusi-Rasi K, Vuori I. Randomised, controlled walking trials in postmenopausal women: the minimum dose to improve aerobic fitness? Br J Sports Med. 2002;36:189-94.

26. Ross R, Janssen I. Physical activity, total and regional obesity: dose-response considerations. Med. Sci. Sports Exerc., 2001;33: S521–S527.

27. Pescatello LS, Franklin BA, Fagard R, Farquhar WB, Kelley GA, Ray CA; American College of Sports Medicine.American College of Sports Medicine position stand. Exercise and hypertension. Med Sci Sports Exerc. 2004;36:533-53.

28. Voorrips LE, Meijers JH, Sol P, Seidell JC, van Staveren WA. History of body weight and physical activity of elderly women differing in current physical activity. Int J Obes Relat Metab Disord. 1992;16:199-205.

29. Ching PL, Willett WC, Rimm EB, Colditz GA, Gortmaker SL, Stampfer MJ. Activity level and risk of overweight in male health professionals. Am J Public Health. 1996;86:25-30.

30 Ainsworth BE, Haskell WL, Leon AS, Jacobs DR Jr, Montoye HJ, Sallis JF, Paffenbarger RS Jr. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc. 1993;25:71-80.

31. Williams PT. Health effects resulting from exercise versus those from body fat loss. Med Sci Sports Exerc. 2001;33:S611-21.

32. Fuller WA. Measurement error Models. 1987 John Wiley & Sons. New York. pp 1-439

33. Williams PT, Wood PD. The effects of changing exercise levels on weight and age-related weight gain. Int J Obes (Lond). 2005 Nov 29; [Epub ahead of print]

Table 1. Regression estimates of the contribution of self-selection to the decline in adiposity measures with walking volume observed cross-sectionally in women.

	Regression coefficients		Tanget slope (1st derivitive)	Estimated effect of a one km increase in weekly walking distance, starting at:		
	km/wk	km/wk^2		0 km/wk	20 km/wk	40 km/wk
BMI						
Starting	-0.0787 (0.0080)‡	0.0009 (0.0001)‡	-0.0787 +0.0017km	-0.0787	-0.0444	-0.0100
Current	-0.1740 (0.0070)‡	0.0017 (0.0001)‡	-0.1740 +0.0034km	-0.1740	-0.1069	-0.0397
Difference	-0.0953 (0.0049)‡	0.0008 (0.0001)‡	-0.0953 +0.0016km	-0.0953	-0.0625	-0.0297
%				45.25	41.55	25.30
Waist						
Starting	-0.0820 (0.0176)‡	0.0009 (0.0003)§	-0.0820 +0.0018km	-0.0820	-0.0453	-0.0085
Current	-0.3717 (0.0206)‡	0.0035 (0.0003)‡	-0.3717 +0.0071km	-0.3717	-0.2300	-0.0882
Difference	-0.2897 (0.0180)‡	0.0026 (0.0003)‡	-0.2897 +0.0052km	-0.2897	-0.1847	-0.0798
%				22.07	19.68	9.62
Hip						
Starting	-0.1980 (0.0243)‡	0.0023 (0.0004)‡	-0.1980 +0.0045km	-0.1980	-0.1078	-0.0175
Current	-0.3019 (0.0191)‡	0.0028 (0.0003)‡	-0.3019 +0.0057km	-0.3019	-0.1888	-0.0757
Difference	-0.1039 (0.0207)‡	0.0006 (0.0003)	-0.1039 +0.0011km	-0.1039	-0.0810	-0.0582
%				65.59	57.09	23.15
Chest						
Starting	-0.1209 (0.0231)‡	0.0015 (0.0004)‡	-0.1209 +0.0030km	-0.1209	-0.0607	-0.0006
Current	-0.2044 (0.0153)‡	0.0019 (0.0002)‡	-0.2044 +0.0039km	-0.2044	-0.1270	-0.0496
Difference	-0.0835 (0.0196)‡	0.0004 (0.0003)	-0.0835 +0.0009km	-0.0835	-0.0663	-0.0490
%				59.14	47.82	1.16
Bra cup						
Starting	-0.0034 (0.0016)*	0.0000 (0.0000)	-0.0034 +0.0001km	-0.0034	-0.0016	0.0001
Current	-0.0112 (0.0016)‡	0.0001 (0.0000)§	-0.0112 +0.0001km	-0.0112	-0.0076	-0.0039
Difference	-0.0078 (0.0010)‡	0.0000 (0.0000)†	-0.0078 +0.0000km	-0.0078	-0.0059	-0.0040

%				30.22	21.69	<0

Significance levels from standard polynomial regression are given by * p<0.05; † p<0.01; § p<0.001; and ‡ p<0.0001. Note: the slope for □□□□□□□□□□□□□□$BMI_{current}$-$BMI_{starting}$) is equal to the difference between the slopes for $BMI_{current}$ and $BMI_{starting}$.

Table 2. Regression estimates of the contribution of self-selection to the decline in adiposity measures with walking intensity observed cross-sectionally in women.

	Regression coefficients		Tanget slope (1st derivitive)	Estimated effect of a one m/s increase in walking speed, starting at:		
	m/s	m/s^2		1.2 m/s	1.8 m/s	2.4 km/wk
BMI						
Starting	-5.5054 (0.3948)‡	0.8705 (0.0998)‡	-5.5054 +1.7409m/s	-3.4162	-2.3716	-1.3270
Current	-8.3605 (0.3435)‡	1.4935 (0.0869)‡	-8.3605 +2.9869m/s	-4.7761	-2.9840	-1.1917
Difference	-2.8551 (0.2449)‡	0.6230 (0.0619)‡	-2.8551 +1.246m/s	-1.3599	-0.6123	0.1353
%				71.53	79.48	>100
Waist						
Starting	-9.0441 (0.9127)‡	1.5806 (0.2296)‡	-9.0441 +3.1612m/s	-5.2507	-3.3540	-1.4572
Current	-20.7640 (1.0585)‡	3.7471 (0.2663)‡	-20.7639 +7.4942m/s	-11.7710	-7.2743	-2.7779
Difference	-11.7198 (0.9398)‡	2.1665 (0.2365)‡	-11.7198 +4.3331m/s	-6.5202	-3.9203	-1.3206
%				44.61	46.11	52.46
Hip						
Starting	-15.7856 (1.2433)‡	2.7458 (0.3125)‡	-15.7856 +5.4916m/s	-9.1957	-5.9008	-2.6058
Current	-14.8500 (0.9804)‡	2.4882 (0.2464)‡	-14.85 +4.9765m/s	-8.8783	-5.8923	-2.9066
Difference	0.9357 (1.0733)	-0.2576 (0.2698)	0.9357 - 0.5151m/s	0.3175	0.0085	-0.3008
%				>100	>100	89.65
Chest						
Starting	-10.1535 (1.1799)‡	1.5413 (0.2976)‡	-10.1535 +3.0827m/s	-6.4544	-4.6048	-2.7553
Current	-11.3859 (0.7841)‡	2.0150 (0.1977)‡	-11.3859 +4.0299m/s	-6.5499	-4.1320	-1.7139
Difference	-1.2325 (1.0104)	0.4736 (0.2548)	-1.2325 +0.9473m/s	-0.0959	0.4726	1.0408
%				98.54	>100	>100
Bra cup						
Starting	-0.5040 (0.0866)‡	0.0875 (0.0219) ‡	-0.5040 +0.175m/s	-0.2940	-0.1890	-0.0840
Current	-0.8147 (0.0842)‡	0.1466 (0.0213) ‡	-0.8147 +0.2932m/s	-0.4629	-0.2870	-0.1110
Difference	-0.3107 (0.0543)‡	0.0591 (0.0137) ‡	-0.3107 +0.1181m/s	-0.1689	-0.0980	-0.0270

%				63.52	65.87	75.66
Significance levels from standard polynomial regression are given by * $p<0.05$; † $p<0.01$; § $p<0.001$; and ‡ $p<0.0001$.						

Table 3. Regression estimates of the contribution of self-selection to the decline in adiposity measures with walking volume observed cross-sectionally in men.

	Regression coefficients		Tanget slope (1st derivitive)	Estimated effect of a one km increase in weekly walking distance, starting at:		
	km/wk	km/wk²		0 km/wk	20 km/wk	40 km/wk
BMI						
Starting	-0.0071 (0.0134)	0.0000 (0.0002)	-0.0071	-0.0071	-0.0071	-0.0071
Current	-0.0679 (0.0115)‡	0.0004 (0.0002)*	-0.0679 +0.0008km	-0.0679	-0.0511	-0.0343
Difference	-0.0608 (0.0082)‡	0.0004 (0.0002)§	-0.0608 +0.0008km	-0.0608	-0.0440	-0.0272
%				10.39	13.89	20.82
Waist						
Starting	0.0147 (0.0416)	0.0002 (0.0006)	0.0147 +0.0004km	0.0147	0.0223	0.0300
Current	-0.1411 (0.0356)‡	0.0007 (0.0005)	-0.1411 +0.0015km	-0.1411	-0.1112	-0.0813
Difference	-0.1558 (0.0372)‡	0.0006 (0.0006)	-0.1558 +0.0011km	-0.1558	-0.1336	-0.1113
%				<0	<0	<0

Significance levels from standard polynomial regression are given by * $p<0.05$; † $p<0.01$; § $p<0.001$; and ‡ $p<0.0001$.

Table 4. Regression estimates of the contribution of self-selection to the decline in adiposity measures with walking intensity observed cross-sectionally in men.

	Regression coefficients		Tanget slope (1st derivitive)	Estimated effect of a one m/s increase in walking speed, starting at:		
	m/s	m/s^2		1.2m/s	1.8 m/s	2.4 km/wk
BMI						
Starting	-1.1956 (0.6138) *	0.0568 (0.1468)	-1.1956 +0.1136m/s	-1.0366	-0.9911	-0.9457
Current	-3.4215 (0.5239)‡	0.5774 (0.1253)‡	-3.4215 +1.1548m/s	-1.8048	-1.3429	-0.8809
Difference	-2.2259 (0.3860)‡	0.5206 (0.0923)‡	-2.2259 +1.0412m/s	-0.7682	-0.3517	0.0648
%				52.03	73.81	>100
Waist						
Starting	3.3959 (1.9571)*	-1.0531 (0.4673)*	3.9359 - 2.1063m/s	0.8685	-0.3953	-1.6590
Current	-6.4058 (1.6848)‡	0.8056 (0.4023)*	-6.4058 +1.6112m/s	-4.4724	-3.5056	-2.5389
Difference	-10.3418 (1.7548)‡	1.8587 (0.4190)‡	-10.3418 +3.7175m/s	-5.8809	-3.6505	-1.4200
%				<0	11.27	65.34

Significance levels from standard polynomial regression are given by * p<0.05; † p<0.01; § p<0.001; and ‡ p<0.0001.

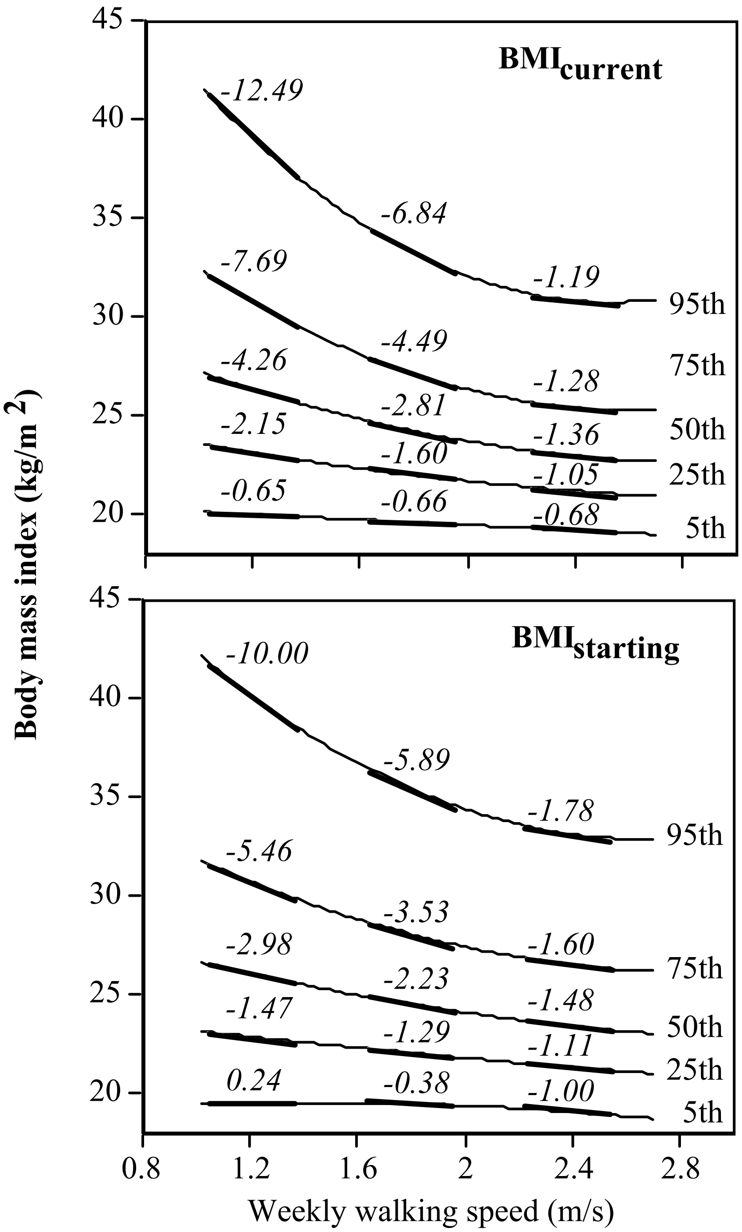

Figure 1. Relationship of self-reported walking speed to different percentiles of women's current and starting BMI (dashed lines), and the slopes (tangents) at 1.2, 1.8 and 2.4 m/s (solid line segments).

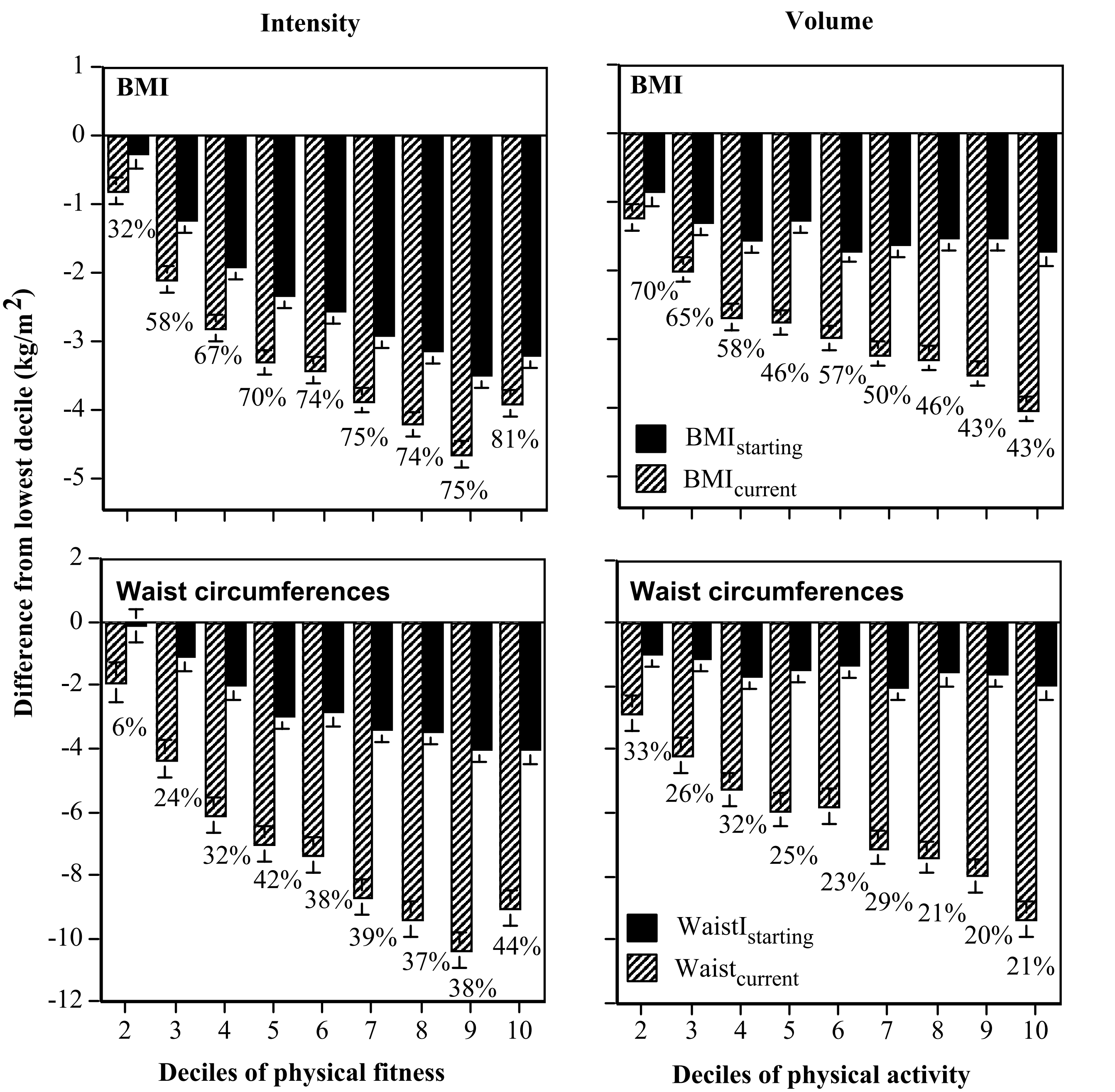

Figure 2 Self-selection of BMI and waist circumference in female walkers. Values below bars represent the proportions accounted for by self-selection, which were estimated as follows: 1) we divided the sample into deciles of walking intensity or walking volume, 2) we calculated the decrease in $BMI_{current}$ at each decile relative to the lowest decile of intensity or volume, and 3) we compared these

decreases to the corresponding differences in starting BMI (specifically their recollection of adiposity when they first began walking 12 or more miles per week). Self-selection was calculated as proportion of the mean reduction in $BMI_{current}$ represented by the mean decrease in starting BMI Values. Negative heights mean faster women who walked longer distances were leaner. All variables age-adjusted.

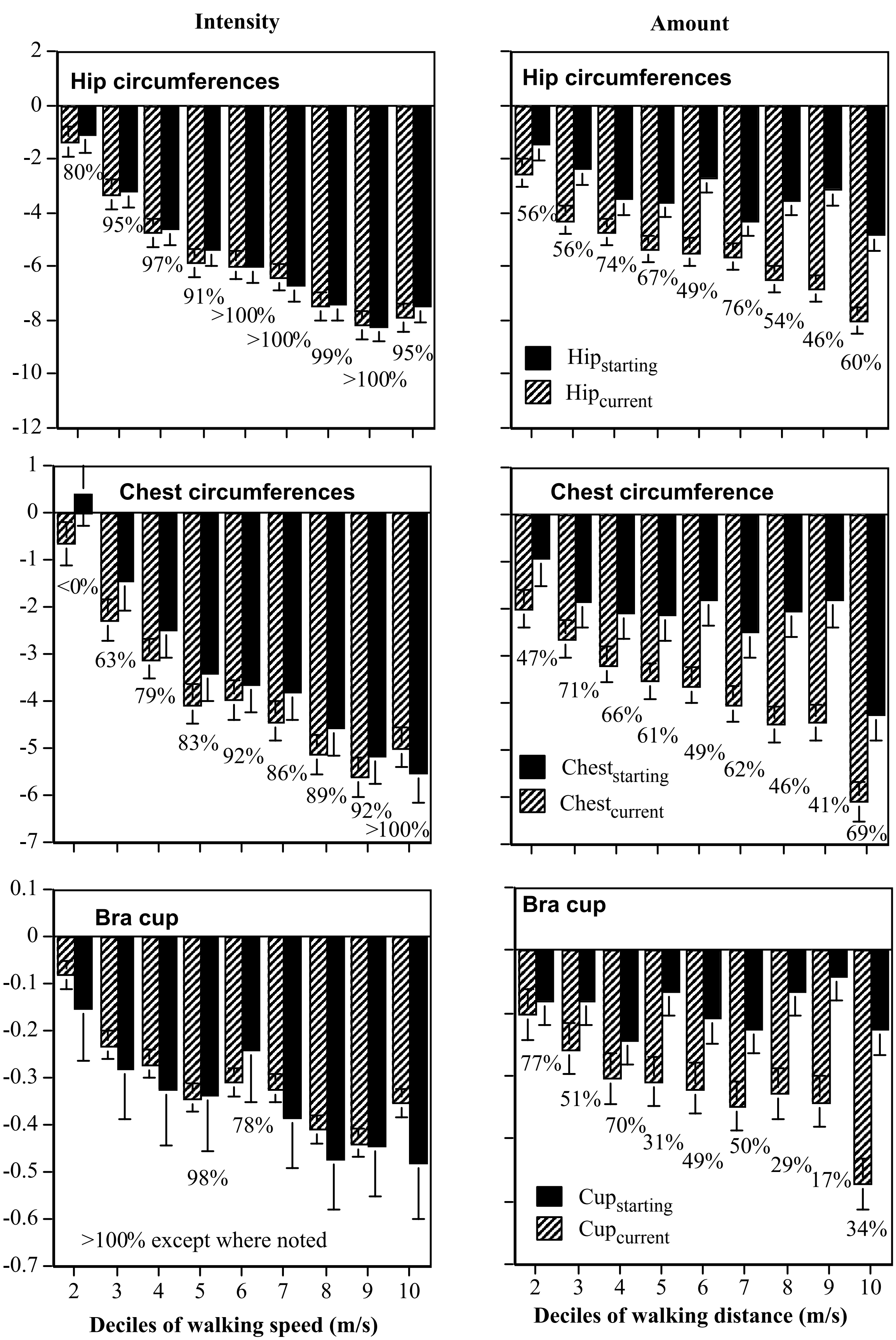
Intensity
Amount
Hip circumferences
80%
95%
97%
91%
>100%
>100%
99%
>100%
95%
Hip circumferences
56%
56%
74%
67%
49%
76%
54%
46%
60%
Hip starting
Hip current
Chest circumferences
<0%
63%
79%
83%
92%
86%
89%
92%
>100%
Chest circumference
47%
71%
66%
61%
49%
62%
46%
41%
69%
Chest starting
Chest current
Bra cup
78%
98%
>100% except where noted
Bra cup
77%
51%
70%
31%
49%
50%
29%
17%
34%
Cup starting
Cup current
Deciles of walking speed (m/s)
Deciles of walking distance (m/s)

Figure 3. Self-selection of hip and chest circumference and bra cup size in female walkers. See caption for Figure 1. Bra-cup sizes were coded on a five-point scale from 1 (A cup), 2 (B cup), 3 (C cup), 4 (D cup), and 5 (E cup or larger).

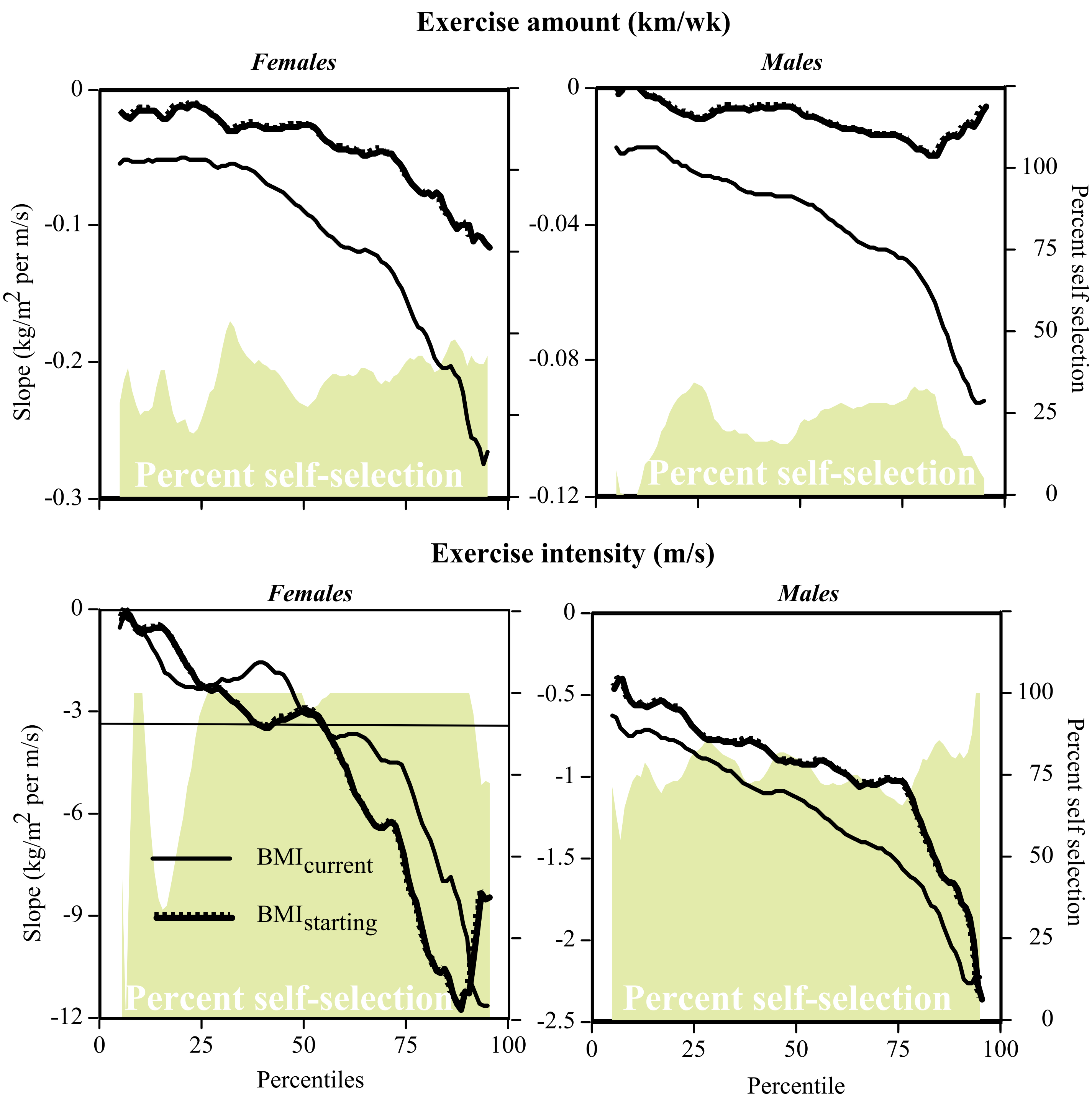

Figure 4. Relationships of the slopes (plotted along Y-axis) of current and starting BMI versus walking dose (distance km/wk) and intensity (velocity, m/s) by percentile of BMI (plotted along X-axis). Shade area represents the percentage of the slope for current BMI

versus dose or intensity attributed to self-selection (scale on the right). Elsewhere, we have shown that the decline in BMI with running distance is greater at the 90th BMI percentile than the 10th percentile, and that the decline becomes progressively greater from the lowest to highest percentile {12-14}. All variables age-adjusted.

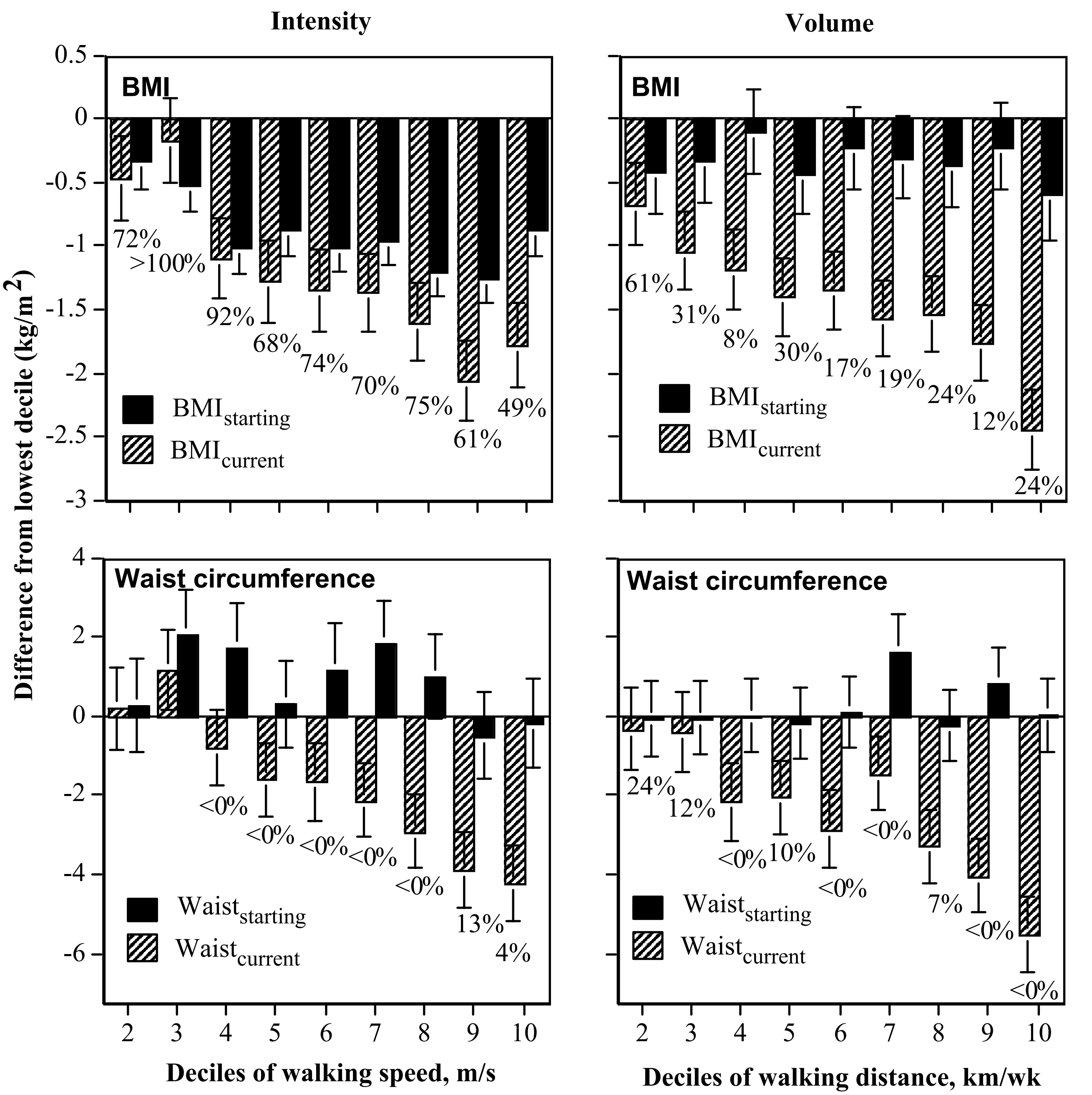

Figure 5. Self-selection of BMI and waist circumference in male walkers. Values below bars represent the proportions accounted for by self-selection. Negative heights mean faster men who walked longer distances were leaner. All variables age-adjusted.