Published in final edited form as:

Am J Health Behav. 2014 March ; 38(2): 254–264. doi:10.5993/AJHB.38.2.11.

Self-reported Activity and Accelerometry in 2 Behavior-maintenance Trials

Meghan M. Senso, MS [Project Manager],

HealthPartners Institute for Education and Research, Minneapolis, MN

Christopher P. Anderson, MPH [Biostatistician Consultant],

HealthPartners Institute for Education and Research, Minneapolis, MN

A. Lauren Crain, PhD [Senior Research Investigator],

HealthPartners Institute for Education and Research, Minneapolis, MN

Nancy E. Sherwood, PhD [Senior Research Investigator], and

HealthPartners Institute for Education and Research, Minneapolis, MN

Brian C. Martinson, PhD [Senior Research Investigator]

HealthPartners Institute for Education and Research, Minneapolis, MN

Abstract

Objectives—To compare accelerometry (MVPA-A) and self-reported activity (MVPA-SR) in activity-maintenance (Keep Active Minnesota; KAM) and weight-loss-maintenance (Keep It Off; KIO) trials.

Methods—Linear regression estimated moderation of study, treatment, or time on MVPA-A and MVPA-SR associations.

Results—MVPA-A was similar between studies (KAM 119 minutes, KIO 112 minutes, $p = .555$), while MVPA-SR differed significantly (KAM 350 minutes, KIO 87 minutes, $p < .0001$). Only study moderated correlation between MVPA-A and MVPA-SR. MVPA-SR better predicted MVPA-A in KIO relative to KAM ($p = .023$).

Conclusions—Results suggest that self-presentation bias may influence validity of self-report in intervention studies with activity as a primary outcome. Researchers should select self-report to assess activity dimensions that objective measures capture poorly.

Keywords

objective monitoring; behavioral intervention; bias

For most of the past 6 decades, self-report has been the primary source for assessing physical activity (PA) and its relationship with health-related outcomes in large-scale epidemiological, public health and behavioral intervention studies.¹ This is due largely to

Correspondence Ms Senso; meghan.m.senso@healthpartners.com.

Conflict of Interest Statement There are no other affiliations that present potential conflicts.

practical constraints associated with objective monitoring of PA (primarily high costs and respondent burden) and despite recognized limitations of such self-reports, most notably, high levels of measurement error.² It is only within the past 15 years that objective measures of PA have become feasible in large samples using relatively inexpensive accelerometers, and only within the past decade or so that the use of such devices has become a common feature of public health surveillance^{3,4} and behavioral intervention studies.⁵⁻⁷

Accelerometry is informative about the frequency, intensity and duration of PA and generally captures activity whether it is purposeful and planned, or merely incidental. Compared to self-report measures, accelerometer-based measures of PA also display stronger associations with PA energy expenditure assessed using doubly-labeled water, with biomarkers, and with health outcomes.⁸⁻¹⁰ Monitor-based objective measures also have their limitations. These include the fact that accelerometry is typically uninformative about the modes or types of activity and their location or context.^{1,9,11} Accelerometry may undercount PA at both low¹² and high¹³ extremes of intensity, and yields incomplete capture of common non-ambulatory types of PA such as biking, primarily upper-extremity activities, and aquatic-based PA.^{14,15} Moreover, a recent study has shown more consistent correlations between self-reported *sedentary behavior* and cardiometabolic risk factors than sedentariness based on accelerometry.¹⁶ Also of concern is the lack of consistency in accelerometry data summary protocols throughout the field making cross-study comparisons challenging.

Such limitations notwithstanding, the state-of-the-art of objective PA monitoring continues to evolve rapidly. With increasing use of pattern recognition techniques for data summarization, and other ongoing improvements in objective monitoring technologies themselves,^{11,17,18} some of these limitations, such as the inability to capture activity mode, may be overcome or mitigated in the near future. Moreover, monitor-based measures have the distinct benefits of being relatively immune to subject reactivity,¹⁹ (though see Motl et al, 2012 for a possible exception²⁰) and other potential biases of self-reported PA, such as recall bias or self-presentation bias. In the context of public health surveillance and behavioral intervention studies, the most potent threat to the validity of self-reported PA appears to be self-presentation bias. Comparisons of accelerometer-based and self-reported PA in general populations are consistent with concerns about over-reporting.^{3,8,10,21,22} Whether consciously or unconsciously, individuals may over-report PA out of recognition that being physically active is a socially approved behavior and sedentariness is often seen as undesirable. Some evidence suggests that individuals scoring higher on the personality trait “social desirability” may be differentially likely to over-report their PA on self-report measures.²³ The potential for self-report bias is particularly concerning in behavioral intervention studies where it is typically not possible to blind either participants or researchers to treatment arm, and participants assigned to treatment or control conditions may consciously or unconsciously report what they expect researchers are hoping to find. It is easy to imagine how such influences could lead to over-reporting of PA in general and differential over-reporting in individuals assigned to study “treatment” arms in particular.

Numerous studies in diverse samples have compared associations between monitor-based objective measures and self-report measures. While some sub-group differences have been

observed, results of these comparisons are consistent in finding only modest correlations ranging from roughly 0.20 to 0.50, averaging around 0.35).^{3,8–10,15,22,24–26} Concerns about bias have led some to suggest objective monitoring of PA may be superior to self-report for public health surveillance and particularly for assessing the impact of behavioral interventions on physical activity.^{22,27} Approaches that combine self-reported PA with more objective measures have also been recommended,^{1,9} though no standard method of doing so has yet emerged. As PA monitors have become more widely available they are increasingly being used in behavioral intervention studies. Most typically, however, objective monitoring has been used as an adjunct to self-reported PA measures, which remain the primary basis upon which behavioral intervention studies of PA initiation or maintenance have been powered and evaluated.^{6,28–33} In considering just how objective and self-reported PA measures might be used in conjunction to increase the validity of outcome measures, one obvious possibility would be to derive adjustment factors to use objective measures to “correct” self-reported PA, similar to those used to correct for bias in self-reported weight. The ability to develop such correction factors is predicated on our knowing more about variability of over-reporting under different conditions.

In this study, we compare accelerometry-based measures of PA to self-reported PA in 2 large-scale behavioral intervention studies conducted by our study team. In these studies we collected both self-report and accelerometer-based measures of PA from randomly selected sub-samples of participants. A number of commonalities across these studies make them an interesting “laboratory” in which to explore the potential for developing correction-factors for self-reported PA. In addition to being conducted by the same study team, both of these intervention studies focused on behavior change *maintenance*, rather than behavior *initiation*, employed the same intervention modality and both had active intervention periods spanning 2 years. Both studies recruited study participants from the same managed care organization in the Minneapolis/St. Paul metropolitan area, and in both studies accelerometry was accomplished using similar protocols and devices. Of interest for the purposes of our comparisons, in the first study, the primary focus of the intervention was on PA maintenance.^{7,33} In the second study, the primary focus of the intervention was on weight-loss maintenance.³⁴ Given concerns about bias in self-report measures, self-presentation bias should arguably be a bigger validity threat in behavioral intervention studies where the primary outcome is PA as opposed to some other outcome (such as weight-loss maintenance). These studies also differed in the average age of participants recruited and in the self-report measures employed (CHAMPS³⁵ and Paffenbarger,³⁶ respectively). Therefore, our aim is to describe the associations between self-reported and objectively measured PA across these populations as well as explore how study (PA vs. weight-loss maintenance trial), time (24-month v. baseline), and randomization (intervention v. control) may moderate these associations.

METHODS

Participant Selection

The PA maintenance trial, “Keep Active Minnesota” (KAM) and the weight-loss maintenance trial, “Keep It Off” (KIO), were conducted by our study team during the

intervals 2004-2007 and 2007-2010, respectively. The recruitment processes for KAM and KIO overall are described in detail elsewhere,^{7,34} but several aspects of eligibility criteria are pertinent to understanding some of our results. Specifically, KAM participants were restricted to ages 50-70 at study enrollment and had to have increased their PA to a minimum of 30 minutes of moderate-to-vigorous PA (MVPA) on at least 2 days a week in the prior year. KIO participants were required to be at least 18 years old and to have lost at least 10% of their body weight in the prior year. In both studies, accelerometry was conducted in sub-samples of enrolled participants who were invited to complete accelerometry. In the KAM study, these participants were selected and invited randomly. In the KIO study, funding for conducting the accelerometry sub-study was not obtained until after recruitment was underway, so all sequential enrollees were invited, until the target number of participants was reached.

Measures

Objectively measured physical activity—In both studies, objective PA was measured via ActiGraph accelerometers (ActiGraph LLC, Pensacola, FL). The 7164 model was used for the KAM study and the GT1M model was used for the KIO study. When comparing these 2 models in adult populations, previous researchers showed no significant differences in activity count output and a 96% agreement in minutes categorized into MET categories.³⁷⁻³⁹ In both studies accelerometers were mailed to participants and initialized to start approximately 8 days before their measurement visit. Participants were asked to wear devices clipped on their right hip for 7 consecutive days. They were instructed to remove the device only when sleeping or during water activities (eg, bathing or swimming). Devices were set to collect data in 60-second epochs. The devices were then collected in person and downloaded. Accelerometry data were included in this analysis if at least 4 valid monitoring days with 10 or more hours of wear time were recorded. Non-wear time was defined as a string of 20 minutes or more of zero-counts, allowing for a 2-minute interruption interval of 100 counts or less. The NHANES cutpoint of 2020 counts per minute was used to estimate minutes MVPA.³ A bout of MVPA was defined as 10 or more consecutive minutes at or above this cutpoint, allowing for a 2-minute interruption interval.

Self-reported physical activity

CHAMPS: In the KAM study, self-reported MVPA was collected using the CHAMPS Activities Questionnaire for Older Adults.³⁵ This instrument asks individuals to estimate how many hours per week were spent doing a range of age-appropriate activities over the past month. As outlined by Stewart et al (2001), adjusted MET values were assigned to each activity. The total durations of activities with a MET value of greater than or equal to 3 were summed to estimate time spent in MVPA. Minutes spent in housework or gardening were excluded from the total time spent in MVPA in response to findings from Heckler et al (2012) which suggest that these items are more prone to over-reporting.²⁵

Paffenbarger physical activity questionnaire—In the KIO study, self-reported PA was collected using the Paffenbarger Physical Activity Questionnaire.³⁶ This instrument asks individuals to name leisure time activities done in the past week and list the duration in which they were done. Activities were later assigned MET values according to the

Compendium of Physical Activities, 2000.⁴⁰ The total durations of all listed activities greater than or equal to 3 METs were then summed to produce an estimate of time spent in MVPA.

Data Analysis

The sample was described in terms of characteristics at baseline using means and standard deviations for age and body mass index (BMI), and frequency counts and proportions for the remaining covariates, which were all categorical variables, such as smoking status or ethnicity. Descriptions were performed both overall and by study. We used t-tests to make between-study comparisons of continuous variables (for instance, age and BMI at baseline). Fisher's exact test was used to compare rates of the remaining baseline descriptors between KAM and KIO participants. For our primary question of interest, we fit 4 separate linear regression models to estimate the association between accelerometer-measured bouts of MPVA (MVPA-A) and self-reported minutes of MVPA (MVPA-SR), and the degree to which the association was moderated by study, treatment and time. The unadjusted associations between MVPA-A and MVPA-SR at baseline and 24 months were estimated using Pearson correlations. Moderation effects from time, treatment and study were assessed by fitting models that looked at MVPA-A as a function of MVPA-SR, the moderator, and the interaction between MVPA-SR and the moderator. Subject-level random effects were estimated in each of the moderation models. Higher order effects were also examined, but significant improvements in fit were not seen. Our final regression model estimated all 2-way interactions between MVPA-SR, time, treatment and study. However, we did not attempt to estimate any 3-way interaction terms, because the analysis lacked the statistical power to detect effect sizes that we were likely to observe. In all of the models, the primary concern was whether the coefficient(s) for the 2-way interaction term(s) were significantly different than zero, which would indicate effect modification. In order to keep our analysis from giving disproportionate weight to outlying values of MVPA-A and MVPA-SR, we trimmed observations that were 3 standard deviations above their respective means. For purposes of making the results of our models of MVPA-A more salient, we present predicted means at 3 levels of MVPA-SR: at 0 minutes (x_0), at the overall mean of MVPA-SR (x_m), and at one standard deviation above the mean of MVPA-SR (x_{m+s}). Analysis was done with PROC MIXED in SAS v9.2 (SAS Institute, Cary, NC, USA), except for graphics, which were produced using R v2.9.2 (R Project, Vienna, Austria).

RESULTS

MVPA-A vs. MVPA-SR

From KAM enrollees, 150 were invited to complete the accelerometry sub-study; 38 declined or were not contacted in time to participate. At baseline, 8 accelerometry records were excluded due to monitor failure or insufficient wear time. In the KIO study, of the 152 invited, 19 declined to complete the sub-study and 23 baseline accelerometry records were excluded due to device failure or insufficient wear time. When combining across study and time point, our sample contained the 384 non-missing observations of MVPA-A ranging from 0 to 665 minutes, with a mean and standard deviation of 122 ± 120 minutes. The 431 non-missing observations of MVPA-SR ranged from 0 to 1260 minutes, with a mean of 220

and standard deviation of 254 minutes. Trimming both MVPA variables at 3 standard deviations above the mean resulted in omission of 6 observations due to abnormally high accelerometry reports, and another 3 observations due to abnormally high self-reports (2 CHAMPS and one Paffenbarger survey responses). After trimming, our sample was comprised of a total of 375 coinciding measurements of MVPA-A and MVPA-SR across time points from 104 KAM participants and 116 KIO participants. Characteristics of the participants by study are shown in table 1. In general, KAM participants were older than KIO participants (by 9.9 yrs., $p < .0001$), more likely to be college educated (74% vs. 58%, $p = .002$), and more-likely to be non-Hispanic white (97% vs. 85%, $p = .002$). Across both assessment times, mean MVPA-A measurements were similar between studies, showing averages of 119 ± 109 minutes and 112 ± 106 minutes for KAM and KIO, respectively ($p = .555$). However, KAM participants had much higher self-reported PA than did participants from KIO (350 ± 258 minutes vs. 87 ± 131 minutes, $p < .0001$).

The intraclass correlation between baseline and 24-month MVPA-A measurements of individual participants was 0.480; both within-subject and between-subject sources of variability are evident. Therefore, Pearson correlations were calculated separately for baseline and 24-months. We observed an $r = 0.291$ ($r^2 = 0.085$) between MVPA-A and MVPA-SR at baseline, and an $r = 0.255$ ($r^2 = 0.065$) between MVPA-A and MVPA-SR at 24-months. In quadrant I of figure 1, individual data points across both baseline and 24-month follow-ups are combined. Note the dashed line, along which all observations would fall if the 2 measurements were always equal; that is, if all variability in MVPA-A was attributable to variability in MVPA-SR, or in statistical terms, $r^2 = 1.0$. This line is also presented in parts II, III, and IV as the y-axis differs across graphs.

Separate Bland-Altman plots of the KIO (figure 2) and KAM (figure 3) study data are included to further describe the associations between MVPA-A and MVPA-SR. The dotted horizontal line in the middle of figure 3 shows that the average difference between self-reported and objectively measured MPVA participants was close to zero among KIO participants. The downward linear trend from zero results from KIO participants who self-reported zero minutes of MVPA but for whom accelerometry registered a non-zero amount of MVPA. In contrast, the average difference between MVPA-SR and MVPA-A was positive among KAM participants (Figure 4), meaning that these participants tended to self-report more MVPA than was measured by accelerometry. The general upward trend in the KAM data provides further evidence of over-reporting of MVPA, in that the difference between MVPA-SR and MVPA-A tended to increase, with self-reported MVPA values higher than measured MVPA values, as their average increased.

Moderation of MVPA-A vs. MVPA-SR by Time

We fit a regression model of MVPA-A as a function of MVPA-SR, time, the interaction between the 2, and a subject-level random-effect used to account for correlation between baseline and 24-month observations of the same individual. From this model, predicted average values for MVPA-A at various levels of self-reported activity for baseline and for 24-month follow-up measurements (of participants from both treatment arms of both studies) are shown in quadrant II of Figure 1. Note that the line of 100% correspondence

appears quite steep in this plot, because the vertical representation of MVPA-A spans a smaller range than the horizontal representation of MVPA-SR. The gray line, representing the relationship of interest at baseline, shows that for every additional hour of self-reported activity, the corresponding increase in accelerometer-measured activity is an estimated 8.9 minutes. Meanwhile, the black line, which represents the change in MVPA-A vs. change in MVPA-SR at 24 months, shows an increase of 5.9 minutes of MVPA-A per each additional hour of self-reported activity. To demonstrate this graphically, dotted lines indicate the predicted values of MVPA-A at the mean value for MVPASR (which is 212 minutes, and referred to as x_m) and one standard deviation above the mean value for MVPA-SR (which is 452 minutes, and referred to as x_{m+s}). In other words, given an average level of self-reported activity, $x_m = 212$ minutes, MVPA-A for measurements taken at baseline is estimated to be 117 minutes, while MVPA-A at 24-months is expected to be 108 minutes. Meanwhile, for MVPA-SR when $x_{m+s} = 452$ minutes, the expected amount of accelerometer-measured activity is 147 minutes at baseline, compared to 136 minutes at 24 months. Thus, the increase of one standard deviation's worth of MVPA-SR, 240 minutes, is associated with a corresponding increase of 30 minutes of MVPA-A at baseline, compared to 28 minutes at the 24-month follow-up. However, when we consider the error around the estimated rates of increase, the change in MVPA-A with MVPA-SR is not significantly different in the baseline measurements compared to those taken at 24 months ($p = .795$).

Moderation of MVPA-A vs. MVPA-SR by Treatment

The results of a separate regression model- MPVA-A as a function of MVPA-SR, treatment group, and the interaction between them, and a subject-specific random effect- are shown in quadrant III of Figure 1. In this plot, the solid gray and black lines are predicted averages that represent all participants randomized to control arms, and all participants randomized to intervention arms, respectively (calculated across studies and follow-up times). For control group participants, predicted MVPA-A values are: 94 minutes at x_0 , 122 minutes at x_m ; and 153 minutes at x_{m+s} (right-most dotted line). Intervention participants' MVPA-A by quantity of MVPA-SR has predicted values of 81 (x_0), 105 (x_m) and 131 (x_{m+s}) minutes, respectively. In other words, in increasing a participant's MVPA-SR from the grand mean to one standard deviation above the mean, we would expect a corresponding increase of 31 minutes of MVPA-A for control participants, while we would expect an increase of 26 minutes in intervention group participants. However, this estimated difference in slopes (0.02 minutes MVPA-A per minute of MVPA-SR) is not statistically significant ($p = .666$).

Moderation of MVPA-A vs. MVPA-SR by Study

Our fourth regression model featured MVPA-A as a function of MVPA-SR, study, and their 2-way interaction term, and person-specific intercept term. The fourth quadrant of figure 1 shows the resultant predicted MVPA-A across the range of MVPA-SR, for KAM participants across treatments and times (black line) and for KIO participants across treatments and times (gray line). At the average value of MVPA-SR x_m , the predicted MVPA-A for KAM participants is 101 minutes, while the predicted MVPA-A for KIO participants is 143 minutes. The gap in expected MVPA-A widens substantially when MVPA-SR is one standard deviation above the mean (x_{m+s}): 130 minutes for KAM participants, versus 207 minutes for those in the KIO group. This represents an estimated

difference of 8.9 minutes of MVPA-A per hour of MVPA-SR between studies, which, given the uncertainty inherent in the estimate ($SE = 0.065$ hrs.), is statistically significant ($p = .023$). In other words, MVPA-A minutes were more closely predicted by MVPA-SR in the KIO study relative to the KAM study.

Moderation of MVPA-A vs. MVPA-SR by Time, Treatment, Study, and Interactions

The final model is an amalgam of the previous 3 models, with MVPA-A minutes expressed as a linear combination of MVPA-SR, time, treatment, and study effects, as well as the 6 possible 2-way interaction terms between predictors (MVPA-SR by study, MVPA-SR by treatment, MVPA-SR by time, study by treatment, study by time, and treatment by time) and a random intercept for each participant. The results of this model are shown in Figure 4. Immediately evident is the marked increase in MVPA-A per minute of MVPA-SR seen in observations from the KIO study (solid and dotted lines) relative to the KAM study (dashed and dot-dashed lines) at all combinations of intervention status and time of measurement. This shows the relative importance of the study by MVPA-SR interaction upon MVPA-A. As was observed in the previous model, this interaction was also statistically significant in this model ($p = .035$). The other 2 interaction terms modifying the slope of MVPA-A with MVPA-SR (MVPA-SR by treatment and MVPA-SR by time) were estimated to be nearly zero ($p = .836$ and $p = .992$, respectively). Note that the line representing the “KAM, Control, Baseline” group falls almost exactly on top of the line representing the “KAM, Control, 24 Mos.” group.

Additionally, the other 3 interaction terms in this model were not statistically different from zero ($p = .892$ for study by time, $p = .444$ study by treatment, and $p = .365$ for treatment by time). Note that, in a graphical sense, the role of these 3 terms is to adjust the y-intercept of a group-specific line on the plot of MVPA-SR vs. MVPA-A (see Figure 4), much in the same way the main effect model terms act. Thus, regardless of the total amount of variability in MVPA-A explained by the possible 2-way interactions between time, treatment and study indicators, the slope of each line representing participants with a given combination of these is unaffected. Rather, the entire line is shifted up or down by a number of MVPA-A minutes, equal to the magnitude of the corresponding term in the regression model.

DISCUSSION

As with other health-related behaviors, physical activity (PA) measurement is nuanced and faces numerous challenges. While large-scale epidemiologic and behavioral intervention studies have traditionally relied on self-reported PA, over the past decade, there has been a shift in favor of using objective measures. This has resulted in part from increased availability of objective PA monitoring, but has also been driven by increasing concern regarding threats to validity posed by widely recognized biases in self-reported measures of PA. The primary purpose of this study was to describe the effects of time, treatment and study on correlations between self-reported and objectively measured PA in 2 behavior-change maintenance trials. Also of importance was our interest in exploring how to more accurately describe PA in future public health surveillance and behavioral intervention research.

Consistent with previous research, we observed only moderate correlations between self-reported and objectively-measured PA in these 2 study populations regardless of time or randomization assignment (baseline $r^2 = 0.085$; 24-month $r^2 = 0.065$). While these were statistically significant ($p < .0001$), it implies that the remaining 91.5% to 93.5% of the variability in accelerometry-measured PA is explained by factors other than self-reported PA. When examining this relationship by time point (baseline vs. follow-up) and randomization (intervention vs. control) we see no significant moderation effects. However, when looking at the moderation by study (KAM vs. KIO), we see a significant interaction, such that minutes of moderate-to-vigorous physical activity as measured by accelerometry (MVPA-A) were more closely predicted by minutes measured by self report (MVPA-SR) in the KIO study relative to the KAM study. This is consistent with the idea that self-presentation bias may be a more potent threat to validity of self-reported PA measures in behavioral intervention studies where the primary outcome is PA as opposed to other outcomes such as weight-loss maintenance. Along with these moderate correlations, participants over-reported minutes spent in MVPA when compared to accelerometer-measured bouts of MVPA regardless of time, treatment, or study. Again this is consistent with previous literature, suggesting the presence of self-presentation bias regardless of time, treatment, or study. However, because there were no significant interactions by randomization over time, it does not appear that there was differential over-reporting within a specific study by time or treatment group.

Although self-report was only moderately correlated with accelerometry and there is evidence of self-presentation bias, we recognize that there are valid reasons *not* to expect a perfect correlation between self-report and objective measures. Among these is the large range of variability in what we consider PA behavior, including variability in type, intensity, duration, and location. Additionally, as compared to other health outcomes such as BMI, PA can be highly variable from week to week. It can also vary greatly in purposefulness, ranging from planned exercise bouts to non-purposeful, incidental movement. The latter proves extremely difficult to capture with self-report, other biases aside. Our analyses have shown, even in 2 very similar populations with similar measurement protocols, over-reporting was highly differential. This between-study variability along with large within-person variability makes the development of a universal correction factor to compensate for over-reporting unrealistic. Given this, it appears that using self-report alone to measure time spent in MVPA is, at best, a stopgap until more cost-effective objective measures become widely available.

Conversely, although it may be unrealistic to develop a universal correction factor or achieve a high correlation between self-reported and measured activity, it may not be a necessity depending on the main outcome of a given study. For example, if the primary focus of study is to measure purposeful or planned activity, such as group exercise for improved social support, then self-reported time spent in these deliberate activities may be adequate. Additionally, until researchers are better able to use objective measures to accurately differentiate between types of activities, supplementing objective activity measurement with self-report is needed to fully describe PA behavior. It is important to remember however, that many self-report measures were originally developed to estimate

time spent in activity rather than activity type. For these measures to be effectively used in conjunction with objective measures, researchers should select self-report measures that focus specifically on measuring types of PA rather than those estimating time.

Some limitations of this study should be kept in mind. First, although the significant between study difference in associations of MVPA-A and MVPA-SR could be due to differential presence of self-presentation bias, we cannot rule out other differences between the 2 studies as possibly contributing factors. These include differences in the average age of participants resulting from eligibility criteria in KAM and KIO (57.2 ± 4.5 years and 47.4 ± 10.7 years, respectively) and different self-report measures in the 2 studies (CHAMPS and Paffenbarger, respectively). More specifically, CHAMPS asks participants to report PA over a typical week in the past month and Paffenbarger asks participants about the previous week. This difference in reporting period could be a competing reason for the differential over-reporting seen in this analysis. Second, while these studies were both large-scale trials, budget limitations precluded conducting accelerometry on the total study samples. And although the processes for inviting enrollees varied across study, (randomly in KAM and sequentially in KIO), there is no reason to believe participants enrolled later in the KIO trial were systematically different than those enrolled earlier. Therefore, this likely does not explain the study moderation effect. Third, while initial willingness to participate in the accelerometry sub-studies was quite good (overall 86%), and balance between the sub-groups of interest was maintained, various data losses resulted in complete accelerometry data for slightly less than 60% of those invited, which reduced our statistical power. Finally, although a standard data summary protocol of defining accelerometry compliance as at least 4 days of 10 or more hours of valid wear time was used, we acknowledge the potential this protocol has for misrepresenting a participant's overall activity level.

Limitations notwithstanding, this study makes a significant contribution to knowledge by highlighting the importance of using both objective and self-report measures in combination to accurately describe PA. As shown, there is a significant difference in the correlation between self-reported and objectively-measured PA by study even with similar interventions, measurement protocols, populations, and study purposes. Given this, self-report does not appear to be sufficient for quantifying amount of PA and a simple correction factor may be challenging, if not impossible, given the large and multi-dimensional variability in PA itself. Although objective measures may be necessary to quantify amount of activity, self-report measures can offer value in describing the aspects of PA that accelerometry cannot currently capture. Moving forward, we see multiple routes to improve measurement of PA for behavioral interventions and public health surveillance. First, it seems highly likely that objective measures of PA will become increasingly central to quantifying PA in both arenas. Yet, as objective measurement itself is evolving towards the use of multiple monitors and assessment of multiple dimensions (eg, posture, heart-rate, respiration), continued work is needed to reduce the monetary and participant burden costs incurred. Second, researchers should reorient their thinking about self-report measures to move away from self-report for quantifying amount of activity and instead use self-reports for purposes such as describing type or purposefulness of PA. When deciding which PA measures to use it is also important to recognize the relative ability of each measure to capture specific components, as there is considerable variability in performance across self-

report measures and across objective measures. Third, further work seems warranted to develop or evolve self-report measures that are more appropriate for use as adjunct measures to objective PA monitoring efforts.

Human Subjects Statement

Informed consent was obtained from participants in each study prior to participation and all study protocols were approved by an institutional review board.

Acknowledgments

Accelerometer data were processed and summarized using a Microsoft Excel macro tool developed and provided by Dr Stewart G. Trost of The University of Queensland, Australia.

The Keep Active Minnesota Study was funded by a grant from the National Institute on Aging (R01 AG023410). The Keep It Off Study was funded by a grant from the National Cancer Institute (R01 CA128211). Both studies are also supported by National Institute of Diabetes and Digestive and Kidney Diseases (P30DK050456 and P30DK092924) and are registered at www.clinicaltrials.gov (No. NCT00283452 and No. NCT00702455, respectively).

All authors receive grant funding to their institution from the National Institutes of Health. Dr. Sherwood also receives grant funding to her institution through Jenny Craig and the University of California-San Diego. Dr. Martinson consults with the University of Illinois at Urbana-Champaign, Michigan State University, and the Educational Development Center in Boston, MA and has received payment or travel reimbursement from the Hospital for Sick Kids, Toronto, ON, Canada, the Fred Hutchinson Cancer Research Center, the National Academy of Engineering, the University of Wisconsin-Madison, the Council of Graduate Schools-Project on Scholarly Integrity, and the CTSA Clinical Research Ethics Key Function Committee.

REFERENCES

1. Haskell WL. Physical activity by self-report: a brief history and future issues. *J Phys Act Health*. 2012; 9(suppl):S5–S10. [PubMed: 22287448]
2. Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. *Am J Epidemiol*. 2007; 166(7):832–840. [PubMed: 17670910]
3. Troiano RP, Berrigan D, Dodd KW, et al. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008; 40(1):181–188. [PubMed: 18091006]
4. Colley RC, Garriguet D, Janssen I, et al. Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. *Health Rep*. 2011; 22(1):7–14. [PubMed: 21510585]
5. Opdenacker J, Boen F, Coorevits N, Delecluse C. Effectiveness of a lifestyle intervention and a structured exercise intervention in older adults. *Prev Med*. 2008; 46(6):518–524. [PubMed: 18405960]
6. King AC, Friedman R, Marcus B, et al. Ongoing physical activity advice by humans versus computers: The Community Health Advice by Telephone (CHAT) Trial. *Health Psychol*. 2007; 26(6):718–727. [PubMed: 18020844]
7. Sherwood NE, Martinson BC, Crain AL, et al. A new approach to physical activity maintenance: rationale, design, and baseline data from the Keep Active Minnesota trial. *BMC Geriatr*. 2008; 8(1): 17. [PubMed: 18655709]
8. Atienza AA, Moser RP, Perna F, et al. Self-reported and objectively-measured activity related to biomarkers using NHANES. *Med Sci Sports Exerc*. 2011; 43(5):815–821. [PubMed: 20962693]
9. Harris TJ, Owen CG, Victor CR, et al. A comparison of questionnaire, accelerometer, and pedometer: measures in older people. *Med Sci Sports Exerc*. 2009; 41(7):1392–1402. [PubMed: 19516162]
10. Colbert LH, Matthews CE, Havighurst TC, et al. Comparative validity of physical activity measures in older adults. *Med Sci Sports Exerc*. 2011; 43(5):867–876. [PubMed: 20881882]

11. Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. *Med Sci Sports Exerc.* 2006; 38(9):1626–1634. [PubMed: 16960524]
12. Le Masurier GC, Tudor-Locke C. Comparison of pedometer and accelerometer accuracy under controlled conditions. *Med Sci Sports Exerc.* 2003; 35(5):867–871. [PubMed: 12750599]
13. Rowlands AV, Stone MR, Eston RG. Influence of speed and step frequency during walking and running on motion sensor output. *Med Sci Sports Exerc.* 2007; 39(4):716–727. [PubMed: 17414811]
14. Miller R, Brown W, Tudor-Locke C. But what about swimming and cycling? How to “count” non-ambulatory activity when using pedometers to assess physical activity. *J Phys Act Health.* 2006; 3(3):257–266.
15. Prince SA, Adamo KB, Hamel ME, et al. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act.* 2008; 5(1): 56. [PubMed: 18990237]
16. Stamatakis E, Hamer M, Tilling K, Lawlor DA. Sedentary time in relation to cardio-metabolic risk factors: differential associations for self-report vs accelerometry in working age adults. *Int J Epidemiol.* 2012; 41(5):1328–1337. [PubMed: 22634868]
17. Strath SJ, Brage S, Ekelund U. Integration of physiological and accelerometer data to improve physical activity assessment. *Med Sci Sports Exerc.* 2005; 37(11):S563–S571. [PubMed: 16294119]
18. Ward DS, Evenson KR, Vaughn A, et al. Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc.* 2005; 37(11):S582–S588. [PubMed: 16294121]
19. Behrens TK, Dinger MK. Motion sensor reactivity in physically active young adults. *Res Q Exerc Sport.* 2007; 78(2):1–8. [PubMed: 17479568]
20. Motl RW, McAuley E, Dlugonski D. Reactivity in baseline accelerometer data from a physical activity behavioral intervention. *Health Psychol.* 2012; 31(2):172–175. [PubMed: 22023436]
21. Emaus A, Degerstrom J, Wilsgaard T, et al. Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromso study. *Scand J Public Health.* 2010; 38(5):105–118. [PubMed: 21062845]
22. Ham SA, Ainsworth BE. Disparities in data on Healthy People 2010 physical activity objectives collected by accelerometry and self-report. *Am J Public Health.* 2010; 100(S1):S263–S268. [PubMed: 20147669]
23. Adams SA, Matthews CE, Ebbeling CB, et al. The effect of social desirability and social approval on self-reports of physical activity. *Am J Epidemiol.* 2005; 161(4):389–398. [PubMed: 15692083]
24. Friedenreich CM, Courneya KS, Neilson HK, et al. Reliability and validity of the Past Year Total Physical Activity Questionnaire. *Am J Epidemiol.* 2006; 163(10):959–970. [PubMed: 16524954]
25. Hekler EB, Buman MP, Haskell WL, et al. Reliability and validity of CHAMPS self-reported sedentary-to-vigorous intensity physical activity in older adults. *J Phys Act Health.* 2012; 9(2): 225–236. [PubMed: 22368222]
26. Pruitt LA, Glynn NW, King AC, et al. Use of accelerometry to measure physical activity in older adults at risk for mobility disability. *J Aging Phys Act.* 2008; 16(4):416–434. [PubMed: 19033603]
27. Berke EM, Choudhury T, Ali S, Rabbi M. Objective measurement of sociability and activity: mobile sensing in the community. *Ann Fam Med.* 2011; 9(4):344–350. [PubMed: 21747106]
28. Napolitano MA, Papandonatos GD, Lewis BA, et al. Mediators of physical activity behavior change: a multivariate approach. *Health Psychol.* 2008; 27(4):409–418. [PubMed: 18642998]
29. Hammerback K, Felias-Christensen G, Phelan EA. Evaluation of a telephone-based physical activity promotion program for disadvantaged older adults. *Prev Chronic Dis.* 2012; 9:E62. [PubMed: 22360873]
30. Teri L, McCurry SM, Logsdon RG, et al. A randomized controlled clinical trial of the Seattle protocol for activity in older adults. *J Am Geriatr Soc.* 2011; 59(7):1188–1196. [PubMed: 21718259]

31. Castro CM, Pruitt LA, Buman MP, King AC. Physical activity program delivery by professionals versus volunteers: the TEAM randomized trial. *Health Psychol.* 2011; 30(3):285–294. [PubMed: 21553972]
32. Van Stralen MM, De Vries H, Bolman C, et al. Exploring the efficacy and moderators of two computer-tailored physical activity interventions for older adults: a randomized controlled trial. *Ann Behav Med.* 2010; 39(2):139–150. [PubMed: 20182833]
33. Martinson BC, Sherwood NE, Crain AL, et al. Maintaining physical activity among older adults: 24-month outcomes of the Keep Active Minnesota randomized controlled trial. *Prev Med.* 2010; 51(1):37–44. [PubMed: 20382179]
34. Sherwood NE, Crain AL, Martinson BC, et al. Keep it off: A phone-based intervention for long-term weight-loss maintenance. *Contemp Clin Trials.* 2011; 32(4):551–560. [PubMed: 21453791]
35. Stewart AL, Mills KM, King AC, et al. CHAMPS physical activity questionnaire for older adults: outcomes for interventions. *Med Sci Sports Exerc.* 2001; 33(7):1126–1141. [PubMed: 11445760]
36. Paffenbarger RS, Wing AL, Hyde RT. Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol.* 1978; 108(3):161–175. [PubMed: 707484]
37. Rothney MP, Apker GA, Song Y, Chen KY. Comparing the performance of three generations of ActiGraph accelerometers. *J Appl Physiol.* 2008; 105(4):1091–1097. [PubMed: 18635874]
38. Kozey SL, Staudenmayer JW, Troiano RP, Freedson PS. Comparison of the ActiGraph 7164 and the ActiGraph GT1M during self-paced locomotion. *Med Sci Sports Exerc.* 2010; 42(5):971–976. [PubMed: 19997000]
39. John D, Tyo B, Bassett DR. Comparison of four ActiGraph accelerometers during walking and running. *Med Sci Sports Exerc.* 2010; 42(2):368–374. [PubMed: 19927022]
40. Ainsworth, BE.; Haskell, WL.; Herrmann, SD., et al. The compendium of physical activities tracking guide. Healthy Lifestyles Research Center, College of Nursing & Health Innovation, Arizona State University; Available at: <https://sites.google.com/site/compendiumofphysicalactivities/> [Accessed October 9, 2012]

Figure 1.
Moderation of MVPA-SR vs. MVPA-A by Time, Treatment, or Study.

Figure 2.
Bland-Altman Plot of MVPA-SR and MVPA-A, KIO Study.

Figure 3.
Bland-Altman Plot of MVPA-SR and MVPA-A, KAM Study.

Figure 4.
Moderation of MVPA-A vs. MVPA-SR by Time, Treatment, Study, and Interactions.

Table 1

Participant Characteristics at Baseline.

	KAM	KIO	Overall	p-value
N	104	116	220	
Age, M (SD)	57.2 (4.5)	47.3 (10.7)	52.0 (9.7)	< .001
BMI, M (SD)	27.7 (5.6)	27.9 (4.6)	27.8 (5.1)	.805
Non-Hispanic white, N (%)	101 (97.1)	99 (85.3)	201 (90.9)	.002
Male, N (%)	23 (22.1)	21 (18.1)	44 (20.0)	.502
College Graduates, N (%)	77 (74.0)	67 (57.8)	145 (65.5)	.016
Annual Income >\$75k/yr, N (%)	23 (22.1)	37 (31.9)	60 (27.3)	.130
Current Smokers, n (%)	6 (5.8)	9 (7.8)	15 (6.8)	.603
Married, N (%)	72 (69.2)	72 (62.1)	144 (65.5)	.320