American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2015. This work is written by (a) US Government employee(s) and is in the public domain in the US.

Vol. 181, No. 9
DOI: 10.1093/aje/kwu472
Advance Access publication:
April 8, 2015

Invited Commentary

Invited Commentary: Meta-Physical Activity and the Search for the Truth

Hannah Arem*, Sarah Kozey Keadle, and Charles E. Matthews

* Correspondence to Dr. Hannah Arem, NCI Shady Grove, Division of Cancer Epidemiology and Genetics, 9609 Medical Center Drive, Room 6E314, MSC 9768, Bethesda, MD 20892-9768 (e-mail: Aremhe2@mail.nih.gov).

Initially submitted November 26, 2014; accepted for publication December 3, 2014.

Measurement error in self-reported data from questionnaires is a well-recognized challenge in studies of physical activity and health. In this issue of the *Journal*, Lim et al. (*Am J Epidemiol*. 2015;181(9):648–655) used data from accelerometers in a small measurement study to correct self-reported physical activity data from a larger study of adults from New York City and to develop an error correction model. They showed that correction of measurement error in self-reported physical activity levels strengthened the associations of physical activity with both obesity and diabetes by 30%–50% compared with using the self-reported questionnaire data alone. Thus, Lim et al. demonstrated a method to improve potentially biased estimates of the association between self-reported physical activity and disease. However, as this field develops, we feel it is important to call attention to a sometimes overlooked problem that occurs when comparing these instruments: Questionnaires and accelerometers are often calibrated (i.e., designed) to measure different types of physical activity, and accelerometers are still subject to measurement error. Thus, physical activity estimates corrected with an imperfect accelerometer measurement might over- or undercorrect the strength of the associations. We take this opportunity to further comment on physical activity measurement in epidemiologic studies and the implications for research.

accelerometer; exercise; measurement; physical activity; self-report

Abbreviation: GPAQ, Global Physical Activity Questionnaire; MVPA, moderate to vigorous physical activity.

One of the major challenges in measuring physical activity is that there are few gold standards against which self-reported or accelerometer data can be compared in free-living populations. Direct observation can provide a highly accurate measure of physical activity, but acquiring it is resource intensive and not possible on a large scale. Doubly labeled water can be used to capture all energy expenditure, but it does not record information about activity intensity or context and thus offers limited insight into important behavioral metrics, such as time spent in moderate to vigorous physical activity (MVPA).

The physical activity behaviors measured in epidemiologic studies are often specified by the frequency and duration of activity at different intensity levels (light, moderate, or vigorous) and in different domains (household, workplace, transport, or leisure time activities). In this issue of the *Journal,* Lim et al. (1) used a modified version of the Global Physical Activity Questionnaire (GPAQ) that was designed to capture a broad range of MVPAs in multiple domains. They then assigned metabolic equivalent values to the activities based on the Ainsworth compendium (2). Using the GPAQ, they showed modest inverse associations of physical activity with obesity and diabetes. Lim et al. cautioned that studies testing associations between self-reported physical activity and disease outcomes might underestimate associations, sometimes leading to different conclusions about whether an association exists (3).

Thus, they also used a GT3x accelerometer (Actigraph, Pensacola, Florida) to estimate moderate- and vigorous-intensity activities using the count per minute threshold of 2020 or above derived by the National Health and Nutrition Examination Survey (4). However, laboratory studies in which measured metabolic equivalent values and counts per minute were compared have shown that activities such as household work, gardening, and cycling have counts per minute well below the 2020 threshold (5, 6), and recent studies of free-living studies have also suggested bias in this device-based

MVPA measure (7, 8). It is commonly recognized that swimming and cycling are underestimated by accelerometers, but less commonly recognized is the fact that MVPAs accumulated in the workplace and at home can also be underestimated when using this method to collect data. Importantly, many of the MVPAs that the GPAQ measured involved little ambulatory movement (e.g., lifting heavy loads, digging, sweeping, washing windows). Thus, one way to think about the differences between GPAQ and accelerometer measures is that they are calibrated to measure different types of MVPA, and for this reason trying to equate the 2 measures in the context of measurement error correction becomes complicated.

We agree that the version of the GPAQ used in the study by Lim et al. is subject to error and could benefit from measurement error correction. However, we argue that despite the fact that the accelerometer provides an "objective" estimate of MVPA, rigorous validation studies confirming the accuracy and precision of the method are limited, and the available evidence suggests that this approach might provide a biased estimate of MVPA in free-living adults (7–9). In a previous study, Troiano et al. (10) suggested that self-reported and accelerometer-based measures are not interchangeable, but less attention has been directed toward differences in the behaviors that these instruments are designed to measure.

IMPROVING MEASUREMENT IN PHYSICAL ACTIVITY EPIDEMIOLOGY

Despite current limitations, the field of physical activity epidemiology will benefit from development and implementation of better measurement tools in new studies and use measurement error correction approaches (11). In the accompanying paper, Lim et al. applied accelerometer data derived from a smaller measurement study to a larger population-based study that used only self-reported data on physical activity. They effectively demonstrated the potential of an underutilized analytic design in physical activity epidemiology.

Lim et al. evaluated various measurement error modeling assumptions, one of which mandates independent errors between measurement tools. However, meeting this assumption does not overcome concerns that the accelerometer reference may be biased in capturing MVPA. As the authors noted, "The validity coefficient quantifies the validity of the GPAQ measure relative to the accelerometer-based measure" (1, p. 650). Thus, the validity coefficient is only as good as the reference tool. The accelerometer calibration and processing methods used in the accompanying study were developed based on activities such as walking and running, but they underestimate true physical activity levels for household activities and biking (6). Wacholder et al. (12) explored how measurement errors in an "alloyed gold standard" (the accelerometer in this case) can influence the results derived from regression calibration. The amount of error in the "alloyed" reference measure and the direction and magnitude of correlation with the original measurement tool determine the amount and direction of bias in the corrected estimate of the underlying association.

When Lim et al. used measurement error correction for the obesity and diabetes prevalence ratios, the magnitude of protective effect strengthened drastically; however, given the probable measurement error in their accelerometer-based measures, there is uncertainty in the accuracy of the corrected risk estimates. These corrected estimates must be interpreted with the caveat that the validation tool itself is subject to error in measuring physical activity.

WHAT TOOLS SHOULD BE USED TO MEASURE PHYSICAL ACTIVITY?

Questionnaires alone are still a valuable tool for measuring various physical activity domains and constructs in studies in which it is not possible to implement less error-prone measures, such as short-term recalls or objective measures (13, 14). Questionnaires are inexpensive to administer, can query about specific types of activities and time periods, and have been shown to have construct validity with various disease outcomes. Another advantage of questionnaires is that they provide contextual information (i.e., where and why the behavior takes place), which might be valuable for studying specific disease associations and developing interventions and public policy. For instance, when there were large discrepancies between the measurements from the questionnaires and those from the accelerometers (7% of the subsample in the study by Lim et al.), questionnaires could be reviewed to determine the exact types of activities and plausibility of responses.

One of the sources of error associated with questionnaires is test-retest reliability; another is intra-individual variation in actual behavior over time. These issues may be tackled by administration of a second questionnaire and use of regression calibration to adjust for these sources of measurement error without the need for a reference measurement. Although 7-day accelerometry measures produce reliable point estimates for the specific time period, changes in season, health, and work or other obligations could affect whether the measure captures usual physical activity levels. Future studies could consider repeat physical activity measurements using either tool to evaluate improvements in estimating usual physical activity.

In addition, the device-based field can continue to improve physical activity measurement. Accelerometers have become increasingly accessible for use in recent epidemiologic studies because the cost has gone down (15). A limitation in the developing field of methods to use accelerometer data, which was not fully recognized until recently (16), has been the importance of calibrating monitors to capture activities representative of daily living (beyond walking and running) and then conducting rigorous cross-validation studies in independent study samples of free-living persons to quantify the accuracy and precision of the method. Accelerometer calibration methods are improving with advances in technology and processing (e.g., machine-based learning techniques), but the expansion of monitor attachment sites on the body (hip, wrist, etc.) could yield different results. Although these advances in accelerometer technology have currently outpaced our ability to translate the voluminous data into meaningful behavioral metrics, newer methods should help reduce the gap between self-reported and accelerometer-based measures of physical activity behavior and facilitate use of monitors in studies such as the accompanying study (1).

SUMMARY OF RECOMMENDATIONS

Because physical activity is a complex behavior, there is no "one-size-fits-all" approach to data collection and analysis. Each of the measurement tools for physical activity in current use uniquely contributes to our understanding of physical activity in free-living settings. Although some of the challenges of these tools are inherent to the tool itself, other challenges, such as processing and analytic methods, can be improved with more sophisticated techniques. Researchers should consider specific capabilities of accelerometer models, the associated calibration studies, and the availability of cross-validation data from free-living populations. During analysis, careful consideration should be given to processing methods in order to arrive at the best estimation of "true" physical activity levels.

There is a growing body of literature about the application of measurement error modeling and error correction in physical activity epidemiology (9, 17), and Lim et al. are among the first to apply these methods in the context of a population-based complex survey design (1). The authors should be commended for their efforts, and we hope that wider application of this approach will further strengthen the evidence base for physical activity and health.

ACKNOWLEDGMENTS

Author affiliations: Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland (Hannah Arem, Sarah Kozey Keadle, Charles E. Matthews).

Conflict of interest: none declared.

REFERENCES

1. Lim S, Wyker B, Bartley K, et al. Measurement error of self-reported physical activity levels in New York City: assessment and correction. *Am J Epidemiol*. 2015;181(9): 648–655.
2. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc*. 2000;32(9 suppl):S498–S504.
3. Celis-Morales CA, Perez-Bravo F, Ibañez L, et al. Objective vs. self-reported physical activity and sedentary time: effects of measurement method on relationships with risk biomarkers. *PLoS One*. 2012;7(5):e36345.
4. 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.
5. Kozey SL, Lyden K, Howe CA, et al. Accelerometer output and MET values of common physical activities. *Med Sci Sports Exerc*. 2010;42(9):1776–1784.
6. Lyden K, Kozey SL, Staudenmeyer JW, et al. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. *Eur J Appl Physiol*. 2011;111(2):187–201.
7. Crouter SE, DellaValle DM, Haas JD, et al. Validity of ActiGraph 2-regression model, Matthews cut-points, and NHANES cut-points for assessing free-living physical activity. *J Phys Act Health*. 2013;10(4):504–514.
8. Welk GJ, McClain JJ, Eisenmann JC, et al. Field validation of the MTI Actigraph and BodyMedia armband monitor using the IDEEA monitor. *Obesity (Silver Spring)*. 2007;15(4): 918–928.
9. Matthews CE, Keadle SK, Sampson J, et al. Validation of a previous-day recall measure of active and sedentary behaviors. *Med Sci Sports Exerc*. 2013;45(8):1629–1638.
10. Troiano RP, McClain JJ, Brychta RJ, et al. Evolution of accelerometer methods for physical activity research. *Br J Sports Med*. 2014;48(13):1019–1023.
11. Matthews CE, Moore SC, George SM, et al. Improving self-reports of active and sedentary behaviors in large epidemiologic studies. *Exerc Sport Sci Rev*. 2012;40(3): 118–126.
12. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol*. 1993;137(11): 1251–1258.
13. Matthew CE. Calibration of accelerometer output for adults. *Med Sci Sports Exerc*. 2005;37(11 suppl):S512–S522.
14. Welk GJ, Kim Y, Stanfill B, et al. Validity of 24-h physical activity recall: physical activity measurement survey. *Med Sci Sports Exerc*. 2014;46(10):2014–2024.
15. Lee IM, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *Br J Sports Med*. 2014;48(3):197–201.
16. Lyden K, Keadle SK, Staudenmayer J, et al. A method to estimate free-living active and sedentary behavior from an accelerometer. *Med Sci Sports Exerc*. 2014;46(2):386–397.
17. Nusser SM, Beyler NK, Welk GJ, et al. Modeling errors in physical activity recall data. *J Phys Act Health*. 2012;9(suppl 1): S56–S67.