

# Adjusting for Measurement Error and Nonresponse in Physical Activity Surveys: A Simulation Study

Nicholas Beyler<sup>1</sup> and Amy Beyler<sup>2</sup>

Adult Americans are encouraged to engage in at least 150 minutes of moderate to vigorous physical activity (MVPA) each week. National surveys that collect physical activity data to assess whether or not adults adhere to this guideline use self-report questionnaires that are prone to measurement error and nonresponse. Studies have examined the individual effects of each of these error sources on estimators of physical activity, but little is known about the consequences of not adjusting for both error sources. We conducted a simulation study to determine how estimators of adherence to the guideline for adults to engage in 150 minutes of MVPA each week respond to different magnitudes of measurement and nonresponse errors in self-reported physical activity survey data. Estimators that adjust for both measurement and nonresponse errors provide the least amount of bias regardless of the magnitudes of measurement error and nonresponse. In some scenarios, the naïve estimator, which does not adjust for either error source, results in less bias than estimators that adjust for only one error source. To avoid biased physical activity estimates using data collected from self-report questionnaires, researchers should adjust for both measurement error and nonresponse.

*Key words:* Measurement error model; moderate to vigorous physical activity; response propensity model; total survey error.

## 1. Introduction

The U.S. Department of Health and Human Services establishes physical activity guidelines for Americans ([Department of Health and Human Services 2008](#)). A key guideline is for adults to engage in at least 150 minutes of moderate to vigorous physical activity (MVPA) each week. Intensity of physical activity is measured in terms of metabolic equivalents (METs), where an individual at rest is at a baseline level of one MET ([Welk 2002](#)). For adults, moderate physical activity requires three to six METs of intensity, and vigorous activity requires more than six METs of intensity ([Ainsworth et al. 2000](#); [Crouter et al. 2006](#); [Troiano et al. 2008](#)).

One key component of physical activity research is determining the percentage of adult Americans who adhere to this guideline, using information collected via self-report

<sup>1</sup> 2M Research Services, LLC, 1775 Eye Street, NW, Suite 1150, Washington, D.C., U.S.A. Email: [nbeyler@2mresearch.com](mailto:nbeyler@2mresearch.com)

<sup>2</sup> Mathematica Policy Research, 1100 1st Street, NE, Washington, D.C., U.S.A. Email: [abeyler@mathematica-mpr.com](mailto:abeyler@mathematica-mpr.com)

**Acknowledgments:** We would like to thank Mark Timms for his help developing the figures presented in the article and our colleagues from Mathematica's Data Science and Statistics Methodology Working Group (DSSMWG) for their helpful comments on the article. We would also like to thank the editors and reviewers from JOS for their thoughtful comments on the article.

surveys or questionnaires that ask individuals about their physical activity or using monitoring devices such as accelerometers. Troiano et al. (2008) estimated that approximately 51 percent of adult Americans engage in at least 150 minutes of MVPA each week, using physical activity questionnaire data collected for the 2003–2004 National Health and Nutrition Examination Survey (NHANES). Alternatively, using accelerometers from NHANES 2003–2004, which tend to underestimate the amount of physical activity in individuals (Matthews 2005), Troiano et al. (2008) estimated that less than five percent engage in at least 150 minutes of MVPA each week. Tucker et al. (2011) estimated that about 60 percent of adult Americans adhere to the same guideline, using NHANES questionnaire data from 2005–2006, and about ten percent adhere to the guideline using accelerometer data from NHANES 2005–2006. A study using data from the National Health and Interview Survey (NHIS) estimated that closer to 42 percent of adults adhere to the guideline of 150 minutes of MVPA each week (Schoenborn and Stommel 2011), whereas a study of data from the Behavioral Risk Factor Surveillance System (BRFSS) found that approximately 65 percent of adults adhere to the same guideline (Loustalot et al. 2009). Table A in the Appendix provides more summative information about these surveys and the physical activity self-report questionnaires they used.

These studies provide estimates of the same parameter (percentage of individuals who adhere to the guideline for minutes of MVPA) with data collected from the same population (adult Americans), yet yield different results. This is due, in part, to nonsampling errors. Unlike sampling errors, which result from the conscious choice to collect data from a sample instead of an entire population, nonsampling errors are unintentional and can exist in a number of forms (Lessler and Kalsbeek 1992). If there is a difference between a reported or measured value and the true, unknown value for a survey outcome, the consequence is a form of nonsampling error known as measurement error (or response error). Measurement error exists in self-reported physical activity data because of cognitive limitations associated with recalling activities, which may result in unintentional misreporting about the frequency and duration of activity (Bassett et al. 2000; Matthews 2002). Social desirability effects may also entice some respondents to overreport their physical activity from the recent past (Adams et al. 2005; Warnecke et al. 1997). In addition, the terminology used in self-report questionnaires may be confusing or unclear, which can result in misreporting (Sallis and Saelens 2000). Self-reporting errors like these often result in much larger estimates of physical activity than those taken from more objective instruments like accelerometers or other monitoring devices and can fundamentally bias estimates of physical activity in adult populations (Ferrari et al. 2007; Nusser et al. 2012; Tooze et al. 2013).

If some sampled participants fail to respond to survey questions and those individuals would have responded in a systematically different way than sampled participants who do respond to survey questions, there is the potential for nonresponse error (or bias). Like many surveys, those that collect physical activity information through self-reporting are prone to nonresponse. National surveys attempt to reduce the effects of nonresponse bias by using weighting adjustments or other adjustment procedures, but it is unclear whether such approaches are actually beneficial, since physical activity data are unavailable for nonrespondents. The limited research available on the relationship between nonresponse

and physical activity self-reporting in national surveys is inconclusive about whether nonresponse leads to biased estimates of physical activity outcomes. Hill et al. (1997) found that initial survey respondents reported more physical activity relative to survey nonrespondents who were part of a responding follow-up subsample. Van Loon et al. (2003) found that a smaller percentage of survey respondents had low physical activity compared to nonrespondents. However, other studies did not find significant biases in physical activity estimates due to nonresponse (Smith and Nutbeam 1990; Vink et al. 2004).

To obtain accurate estimates for the percentage of adult Americans that adhere to physical activity guidelines, appropriate adjustments for both measurement error and nonresponse are prudent. However, to the best of our knowledge, no study currently exists in which a total survey error framework (Groves and Lyberg 2010) is utilized to account for both measurement error and nonresponse in physical activity self-report surveys to see if and when such an approach is superior to one that adjusts for only one error source or neither error source. Our goal, in this article, is to determine what the combined impacts of measurement error and nonresponse are on estimates of physical activity generated from self-report data. Because no study or data currently exist which focus on both of these error sources, simulation procedures are needed to help achieve this goal. We present results from a simulation study which measures the bias from not adjusting for one or more of these nonsampling errors when estimating adherence to the physical activity guideline for adult Americans to engage in at least 150 minutes of MVPA each week. We developed our simulation models using established statistical methods from the literature for analyzing physical activity data. We present results for simulated scenarios that cover a range of situations in which measurement and nonresponse errors may exist in self-reported physical activity data collected from adult Americans. For each scenario, we consider estimators that account and adjust for both measurement and nonresponse errors, measurement error only, nonresponse error only, and neither error source.

## 2. Methods

In this section we describe the procedures used for the simulation study. A more technical description of the simulation procedures is provided in the Appendix. For the simulation, we assume that a random sample of 1,000 individuals is selected from the population of all adult Americans, and sampled participants are asked to provide information about their physical activity using a self-report survey during two separate weeks, randomly selected, over the course of a year. We assume that each sampled individual has some true, unknown amount of time he or she spends engaging in MVPA during a typical week which is different than the amount he or she reports from the self-report survey. This difference is due to measurement error, which we assume to be random (see Equation 1 in the Appendix). We also assume that some sampled individuals will not provide all the necessary information to measure the amount of time they spend engaged in MVPA from the self-report survey. Each sampled individual is assumed to have some propensity to respond to the survey which is a function of his or her true, unknown physical activity level (see Equation 2 in the Appendix). For the simulation, individuals are randomly assigned as either respondents or nonrespondents according to their response propensity. So for

example, if an individual's propensity to respond is 0.75, he or she has a 75 percent probability of being assigned as a respondent.

Using the simulated data for the 1,000 sampled individuals, we estimate the parameter, the percentage of adults in the population that adhere to the physical activity guideline of 150 or more minutes of MVPA per week (which we denote  $\theta$ ) using four different estimators. One estimator accounts and adjusts for both the measurement error and nonresponse (which we denote  $\hat{\theta}_{full}$ ), one accounts for nonresponse only, ignoring the measurement error (which we denote  $\hat{\theta}_{nr}$ ), one accounts for the measurement error only, ignoring the nonresponse (which we denote  $\hat{\theta}_{me}$ ), and one accounts for neither measurement error nor nonresponse (which we denote  $\hat{\theta}_{naive}$ ). (These estimators are defined more explicitly in the Appendix along with other simulation model parameters.) It is important to focus on all four of these estimators so that we understand not only the combined impact of not adjusting for measurement error and nonresponse (which is represented by the naïve estimator) but also what the impact is if adjustments for only one of these error sources is ignored. Once we calculate estimates for these four estimators based on the sample of 1,000 individuals we repeat the simulation process again (randomly selecting another 1,000 individuals) and calculate four new estimates. In total, we run 1,000 simulations, generating 1,000 replicate estimates for each of the four estimators. We calculate the average bias for each estimator by averaging the differences between the 1,000 simulated estimates and the true value being estimated (the percentage of adults that engage in 150 or more minutes of MVPA each week).

As with any simulation, a number of assumptions are made about the parameter values. The parameter values selected for the simulations (which are provided along with the simulation models in the Appendix) were based on research which studied measurement error and nonresponse in data from physical activity self-report surveys (Ferrari et al. 2007; Tooze et al. 2013; Hill et al. 1997). We assumed that the true percentage of adults that adhere to the physical activity guideline of 150 or more minutes of MVPA each week was  $\theta = 25$  percent. We consider this value reasonable since the available self-report-based estimates of this quantity, ranging from 42 to 65 percent (Troiano et al. 2008; Tucker et al. 2011; Schoenborn and Stommel 2011; Loustalot et al. 2009), are likely too high due to reporting biases, and the accelerometer-based estimates of this quantity, ranging from five to ten percent (Troiano et al. 2008; Tucker et al. 2011), are likely too low, since accelerometers are unable to accurately measure all types of activity in free-living conditions (Matthews 2005).

To investigate the combined impacts of measurement error and nonresponse on the estimators, we considered a range of scenarios. We considered nine different measurement error scenarios (ranging from no random measurement error to very large random measurement error). We also considered nine different nonresponse scenarios (ranging from a strong negative relationship between propensity to respond and physical activity to a strong positive relationship between propensity to respond and physical activity). In total, we considered 81 different simulation scenarios based on the combinations of nine measurement error scenarios and nine nonresponse scenarios. In the results section, we conduct a more detailed examination of six of these scenarios summarized in Table 1 and then summarize findings from all 81 scenarios at the end of the section.

Table 1. Model parameter specifications for six simulation scenarios.

Scenario	Empirical values of key model parameters for Models (1) and (2) in the Appendix
Large measurement error and large negative nonresponse bias	$\sigma_e^2 = 8$ and $\alpha_1 = -2$
Large measurement error and no nonresponse bias	$\sigma_e^2 = 8$ and $\alpha_1 = 0$
Large measurement error and large positive nonresponse bias	$\sigma_e^2 = 8$ and $\alpha_1 = 2$
No measurement error and large negative nonresponse bias	$\sigma_e^2 = 0$ and $\alpha_1 = -2$
No measurement error and no nonresponse bias	$\sigma_e^2 = 0$ and $\alpha_1 = 0$
No measurement error and large positive nonresponse bias	$\sigma_e^2 = 0$ and $\alpha_1 = 2$

Note:  $\sigma_e^2$  represents measurement error variability and  $\alpha_1$  represents nonresponse bias. Both terms are explicitly defined in the Appendix.

### 3. Results

The simulation results for the six scenarios given in Table 1 are presented in terms of average biases in Table 2. There are two separate panels in Table 2 that provide results for scenarios with large measurement error (top panel) and no measurement error (bottom panel). Within each panel, three columns represent three different nonresponse bias conditions – negative nonresponse bias, no nonresponse bias, and positive nonresponse bias. Each table cell includes four average bias estimates for the four estimators of  $\theta$ .

#### 3.1. Simulation Scenarios with Large Measurement Error

Average biases for a scenario in which there is large measurement error and negative nonresponse bias are given in the first column of the top panel of Table 2. The estimators

Table 2. Average biases for estimators of  $\theta$  for six simulation scenarios.

Large measurement error ( $\sigma_e^2 = 8$ in Model (1) in the Appendix)		
Negative nonresponse bias ( $\alpha_1 = -2$ in Model (2) in the Appendix)	No nonresponse bias ( $\alpha_1 = 0$ in Model (2) in the Appendix)	Positive nonresponse bias ( $\alpha_1 = 2$ in Model (2) in the Appendix)
$\hat{\theta}_{naive}: 1.0$	$\hat{\theta}_{naive}: 11.5$	$\hat{\theta}_{naive}: 21.5$
$\hat{\theta}_{nr}: 11.5$	$\hat{\theta}_{nr}: 11.5$	$\hat{\theta}_{nr}: 11.6$
$\hat{\theta}_{me}: -18.0$	$\hat{\theta}_{me}: -0.4$	$\hat{\theta}_{me}: 17.1$
$\hat{\theta}_{full}: -1.3$	$\hat{\theta}_{full}: -0.4$	$\hat{\theta}_{full}: 0.3$
No measurement error ( $\sigma_e^2 = 0$ in Model (1))		
Negative nonresponse bias ( $\alpha_1 = -2$ in Model (2))	No nonresponse bias ( $\alpha_1 = 0$ in Model (2))	Positive nonresponse bias ( $\alpha_1 = 2$ in Model (2))
$\hat{\theta}_{naive}: -19.5$	$\hat{\theta}_{naive}: 0.0$	$\hat{\theta}_{naive}: 15.1$
$\hat{\theta}_{nr}: -0.0$	$\hat{\theta}_{nr}: 0.0$	$\hat{\theta}_{nr}: -0.1$
$\hat{\theta}_{me}: -18.5$	$\hat{\theta}_{me}: 0.1$	$\hat{\theta}_{me}: 17.1$
$\hat{\theta}_{full}: -1.0$	$\hat{\theta}_{full}: 0.0$	$\hat{\theta}_{full}: -0.3$

that account for only one error source do not perform well under this scenario. The average bias for  $\hat{\theta}_{nr}$  is 11.5 percentage points. This means that when accounting for nonresponse bias but not for measurement error, under this scenario estimates of  $\theta$  will be around 36 or 37 percent for a true value of  $\theta$  that is 25 percent. The average bias for  $\hat{\theta}_{me}$  is  $-18.0$  percentage points under this same scenario, so that when accounting for only measurement error, estimates of  $\theta$  will be around seven percent (well below the true value of 25 percent). The naïve estimator,  $\hat{\theta}_{naive}$ , has an average bias of 1.0 percentage points and performs as well as the full-adjustment estimator,  $\hat{\theta}_{full}$ , which has an average bias of  $-1.3$  percentage points.

In the scenario where there is large measurement error and no bias due to nonresponse, the measurement error estimator performs well, as does the full-adjustment estimator. Both have average biases that are 1.0 percentage points in absolute value. Neither the nonresponse adjustment estimator nor the naïve estimator perform well. Both have average biases of 11.5 percentage points.

The final scenario in the top panel of [Table 2](#), with large measurement error and positive nonresponse bias, provides the largest absolute bias in the naïve estimator: 21.5 percentage points. This means that, under this scenario, the naïve estimator provides estimates of  $\theta$  close to 50 percent, almost double the true value of 25 percent. Like the naïve estimator, the nonresponse adjustment and measurement error adjustment estimators do not perform well. The full-adjustment estimator does perform well, with an average bias of only 0.3 percentage points.

### 3.2. Simulation Scenarios with no Measurement Error

The bottom panel of [Table 2](#) provides three scenarios where there is no measurement error. As one would expect, under these scenarios the nonresponse adjustment estimator performs as well as the full-adjustment estimator. The naïve and measurement error adjustment estimators do not perform as well, except for the scenario where there is no nonresponse bias. In the scenario with negative nonresponse bias and no measurement error, the average negative bias for the naïve and measurement error adjustment estimators are both about 19 percentage points. So if these estimators were used under this scenario, the estimates would be around six percent, much less than the true parameter value of 25 percent.

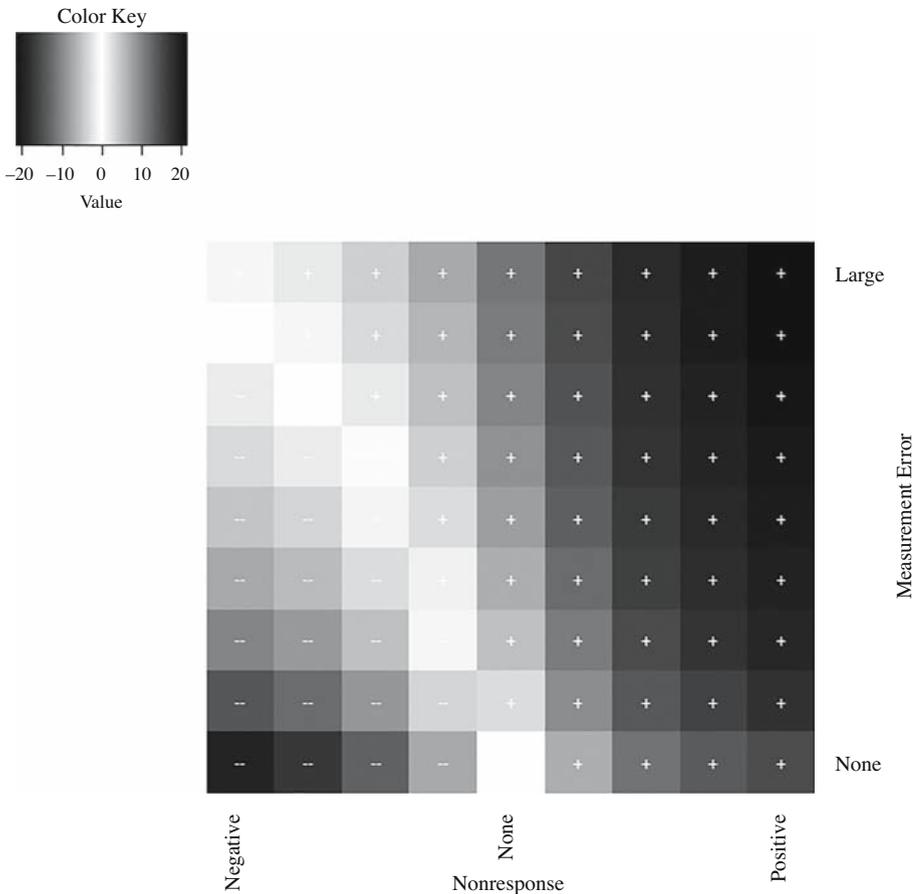
### 3.3. Graphical Representations for Biases in the Naïve and Full Estimators

A more comprehensive picture of the simulation results across all 81 scenarios for the naïve and full estimators is provided in [Figures 1 and 2](#), respectively. In each figure, the x-axis represents the magnitude of the nonresponse bias, which ranges from large negative nonresponse bias (when  $\alpha_1 = -2$ ) to large positive nonresponse bias (when  $\alpha_1 = 2$ ). Along the y-axis, the measurement error ranges from no measurement error (when  $\sigma_e^2 = 0$ ) to large positive measurement error (when  $\sigma_e^2 = 8$ ). The shading in the plot is darker when there is larger bias and the color key in the top left corner of each figure shows the numerical values associated with different shading (little to no shading means that there is little to no bias in the estimator). The “+” and “-” signs indicate whether the bias is positive or negative, respectively. For example, a box that is a dark shade with a “+” sign

means that the estimator was positively biased under the specific scenario the box represents in terms of the measurement error and nonresponse bias scenarios.

In [Figure 1](#), we see that the bias in the naïve estimator varies greatly depending on the magnitudes of the measurement and nonresponse errors. In the bottom left portion of the plot, there is negative bias in the naïve estimator; in the top right portion there tends to be more positive bias in the naïve estimator. There is lighter shading in the middle of the plot, which represent scenarios where there is little to no bias in the naïve estimator. This was also observed in the scenario presented in the top left cell of [Table 2](#), where we saw the naïve estimator perform relatively well.

In comparison to [Figure 1](#), the shading in [Figure 2](#) is consistently much lighter, ranging from only about  $-2$  to  $+2$  percentage points in bias across all scenarios, suggesting that there is always little to no bias in the full estimator regardless of the magnitudes of measurement error and nonresponse bias. This was also reflected in [Table 2](#) where the bias in the full estimator was always about one percentage point or less in absolute value.



*Fig. 1. Average biases in the naïve estimator for simulation scenarios with negative to positive nonresponse bias and no to large measurement error. Darker shades of grey with “-” signs indicate greater negative bias in the naïve estimator; darker shades of grey with “+” signs indicate greater positive bias in the naïve estimator.*

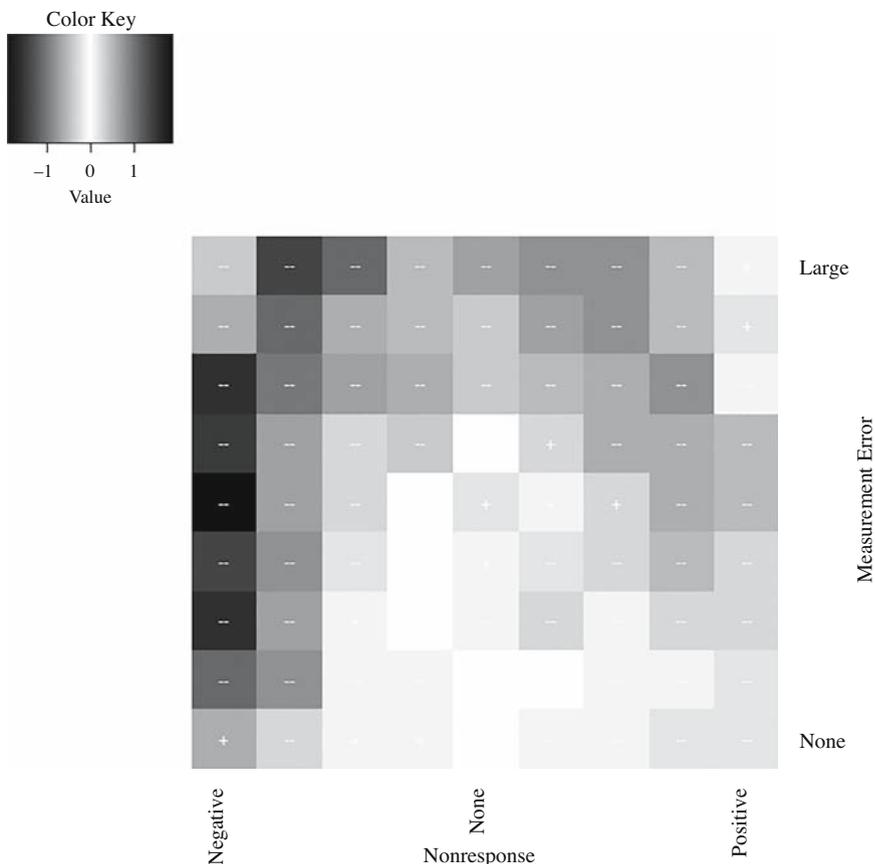


Fig. 2. Average biases in the full estimator for simulation scenarios with negative to positive nonresponse bias and no to large measurement error. Darker shades of grey with “-” signs indicate greater negative bias in the full estimator; darker shades of grey with “+” signs indicate greater positive bias in the full estimator.

#### 4. Discussion

Across all scenarios in our simulation study, which investigate the effects of 81 different combinations of measurement error and nonresponse on estimators of the percentage of adults that adhere to physical activity guidelines, the results suggest that adjusting for both measurement error and nonresponse is preferable to adjusting for one error source or neither error source. The average bias for the full estimator,  $\hat{\theta}_{full}$ , is close to zero regardless of the magnitudes of the measurement error and nonresponse biases (Figure 2). In contrast, the average bias for all other estimators ranges from  $-19.5$  to  $21.5$  percentage points, depending on the scenario being considered (Table 2). There are a few other findings that deserve further discussion.

In scenarios where there is large measurement error and negative nonresponse bias, the naïve estimator, which does not adjust for measurement error or nonresponse, performs better than the estimators that adjust for only one error source (Table 2). This is due to a cancelling-out effect in the error sources. With negative nonresponse bias, more inactive individuals respond than active individuals. But with large measurement error, these

individuals are also misreporting their activity, which creates a positive bias that negates the effect of the negative nonresponse bias in the naïve estimator. The likelihood of scenarios in which there is a negative nonresponse bias in large-scale survey settings may be small. Intuitively, one may actually expect inactive individuals to be less likely to respond to a physical activity survey – not more likely – because they are not as aware of or as interested in their physical activity compared to more active individuals. Moreover, the current, albeit limited, research that looks at relationships between physical activity and nonresponse suggests that the correlation between physical activity and propensity to respond is either positive (Hill et al. 1997; Van Loon et al. 2003) or close to zero (Smith and Nutbeam 1990; Vink et al. 2004).

The more likely simulation scenarios that we could observe in practice may be those in which there is large measurement error and positive nonresponse bias. In these scenarios, the full-adjustment estimator performs well, but the other three estimators do not. The naïve estimator and those that adjust for only one error source produce average biases ranging from 12 to 22 percentage points (Table 2). Therefore, in the more extreme cases, estimates for the percentage of adults that adhere to the physical activity guideline of 150 minutes of MVPA per week that do not adjust for both error sources reach almost 50 percent, whereas the true parameter being estimated is 25 percent. In other words, only one in four adult Americans adheres to the guideline (in our simulation study), whereas estimates that do not properly adjust for measurement error and nonresponse suggest that anywhere between one in three and one in two adult Americans adheres to the guideline. Estimates based on self-report data from national surveys are not far off from 50 percent (Troiano et al. 2008; Tucker et al. 2011; Schoenborn and Stommel 2011; Loustalot et al. 2009). In these studies, weighted estimates are presented that account for the sample design and nonresponse, but nonresponse adjustments are based primarily on demographic characteristics of respondents and nonrespondents, like gender and race, and not on physical activity outcomes. For this reason, the methods considered in this simulation study cannot be directly applied to data from national surveys such as NHANES or NHIS because these surveys do not collect two or more independent measures at different points in time from a subset of individuals. We recommend that future surveys be designed with these specifications to allow for both adjustments. So although it is unclear whether the true value of 25 percent used in the simulations is accurate, the simulations in which we see overestimation of this parameter may be good proxies for what is happening in practice with the self-report data from these national surveys: a combination of positive, nonresponse bias and large measurement error. Respondents in these surveys may tend to be more physically active individuals and may also misreport their activity.

The findings presented in this article are not without their limitations. The findings are based on simulations and should be interpreted as such. The relationship between propensity to respond and physical activity may not be as simplistic as we depict it in Model (2). There may be different response patterns that depend on the gender, age group, racial or ethnic status, or other characteristics of the adult Americans in the population. Moreover, in some subgroups of the population, there may be no association between physical activity and propensity to respond, as some studies suggest (Smith and Nutbeam 1990; Vink et al. 2004). Additional research that sets out to capture the physical activity of

individuals who tend to be nonrespondents in national surveys through other avenues, like field follow up, is needed.

Another caution to note is that the measurement error model (1) used in the simulations assumes that there is no systematic bias in the self-report measurements. This is likely too optimistic an assumption and was used in our simulations to focus, specifically, on random measurement error. Other studies have shown evidence of systematic biases in self-reports of physical activity outcomes (Ferrari et al. 2007; Nusser et al. 2012; Tooze et al. 2013), although they focused on data analyzed from convenience samples that are not representative of all adult Americans. To better understand measurement error properties of self-reported physical activity data in national surveys, multiple measurements from different points in time should be considered for at least a subsample of survey participants in order to distinguish measurement error and other forms of intra-individual variation in the data from actual variability in physical activity across individuals. Such an approach would not require a major change in a survey's design; at a minimum, a single follow-up data collection on a randomly selected subsample of the individuals at a later point in time (after the initial survey data are collected) would be required. As an alternative to this measurement error approach, structural equation models which model multiple indicators of physical activity from national surveys could be considered to better understand the measurement error properties in physical activity measurements.

The study's limitations should not diminish the importance of the findings to the study of physical activity measurement. As part of a comprehensive simulation study, we established 81 different scenarios in which measurement error and nonresponse influence self-reported physical activity data. Although there are a number of other scenarios we could have considered, including those that simulated systematic measurement error or more complex patterns of nonresponse, the ones we focus on provide a good starting point for this line of research focusing on the impacts of multiple sources of nonsampling error on self-reported physical activity outcomes.

## Appendix

### *Technical Documentation for the Simulation Study*

In this Appendix, we provide a more technical presentation of the simulation study set up provided in the methods section.

#### Model Development

For the simulation, we assume that a random sample is selected from the population of all adult Americans, and sampled participants are asked to provide information about their physical activity using a self-report survey. Each individual  $i$  in the sample has some true, unknown amount of time he or she spends engaging in MVPA during a typical week. We define  $T_i$  to be the time spent in MVPA during a typical week for individual  $i$  in minutes. Without loss of generality we assume that in the population, individuals'  $T_i$  are independent and normally distributed with mean  $\mu_T$  and variance  $\sigma_T^2$ . In practice, a log transformation (Ferrari et al. 2007; Tooze et al. 2013) or power transformation (Beyler

et al. 2015) may be required to achieve normality. The main parameter of interest,  $\theta$ , is the percentage of adult Americans that engage in 150 minutes or more MVPA during a typical week. If we knew the value of  $T_i$  for each individual in the sample, we could estimate  $\theta$  as

$$\hat{\theta} = n^{-1} \sum_{i=1}^n I\{T_i \geq 150\},$$

where  $n$  is the number of individuals in the sample and  $I\{T_i \geq 150\}$  is an indicator function that takes a value of 1 if  $T_i$  is greater than or equal to 150, and a value of 0 otherwise. However, in practice,  $T_i$  is unknown and subject to nonsampling errors; therefore, alternative estimation methods must be considered.

We assume that the survey asks each (simulated) individual in the sample to report on his or her physical activity during two weeks randomly selected over the course of a year; from these reports, we obtain two measures for the time he or she spends engaging in MVPA (in minutes) during the course of a typical week. We define  $X_{ij}$  to be the self-report measure for individual  $i$  from week  $j$  ( $j = 1, 2$ ). We assume that  $X_{ij}$  is subject to measurement error and define the relationship between  $X_{ij}$  and  $T_i$  for each individual as

$$X_{ij} = T_i + e_{ij}, \tag{1}$$

where  $e_{ij}$  is a random measurement error term for individual  $i$  from week  $j$ . We assume that the  $e_{ij}$  are independent and normally distributed with mean 0 and variance  $\sigma_e^2$ . We also assume that  $T_i$  and  $e_{ij}$  are independent for all  $i, i'$ , and  $j$ . These assumptions are necessary for model identifiability. Model (1) is often referred to as the classical measurement error model (Carroll et al. 2006) and assumes that  $X_{ij}$  is an unbiased measurement of  $T_i$ . Such an assumption is considered for the purposes of model development for this simulation study but in practice is often not justifiable for self-report measures and more robust model assumptions and model equations should be considered (Ferrari et al. 2007; Nusser et al. 2012; Tooze et al. 2013).

We assume that all individuals sampled will not respond and hence we will not obtain  $X_{ij}$  values from all of our sampled individuals. We assume that each sampled individual  $i$  has a propensity to respond to the survey,  $p_i$ , defined by

$$\log \left\{ \frac{p_i}{1 - p_i} \right\} = \alpha_0 + \alpha_1 T_i, \tag{2}$$

where  $\alpha_0$  and  $\alpha_1$  are fixed intercept and slope coefficients, respectively. In model (2) we assume that for  $\alpha_1 \neq 0$  there is a relationship between an individual's propensity to respond and how physically active he or she is. The full simulation model, accounting for the influence of measurement error and nonresponse on true activity ( $T_i$ ), is represented by model equations (1) and (2).

### Simulation Procedures

In our simulation study, we consider four estimators of  $\theta$ , the true percentage of adult Americans who engage in 150 or more minutes of MVPA each week. We define  $\hat{\theta}_{full}$  to be the estimator of  $\theta$  that accounts and adjusts for both measurement error and nonresponse,  $\hat{\theta}_{nr}$  to be the estimator of  $\theta$  that accounts and adjusts for nonresponse only,  $\hat{\theta}_{me}$  to be the

estimator of  $\theta$  that accounts and adjusts for measurement error only, and  $\hat{\theta}_{naive}$  to be the estimator of  $\theta$  that does not account for either error source. We develop these four estimators in a series of steps:

1. First we randomly generate  $n$  values of  $T_i$  from a normal distribution with mean  $\mu_T$  and variance  $\sigma_T^2$ .
2. For each  $T_i$  we randomly generate two values of  $e_{ij}$  ( $e_{i1}$  and  $e_{i2}$ ) from a normal distribution with mean 0 and variance  $\sigma_e^2$ .
3. Next, we calculate values of  $X_{ij}$  and  $p_i$  from Models (1) and (2), respectively.
4. For each individual  $i$ , we randomly generate  $R_i$  from a Bernoulli distribution with success probability  $p_i$ . If  $R_i$  is 1, individual  $i$  is a respondent; if  $R_i$  is 0, individual  $i$  is a nonrespondent.
5. Because  $p_i$  is not known, in practice, for the simulation we estimate  $p_i$  for all respondents. For individuals with  $R_i = 1$ , we let

$$\hat{p}_i = \frac{\exp\{\alpha_0 + \alpha_1 T_i + \varphi_i\}}{1 + \exp\{\alpha_0 + \alpha_1 T_i + \varphi_i\}},$$

where  $\varphi_i$  is randomly generated from a normal distribution with mean 0 and variance  $\sigma_\varphi^2$ . Each respondent is then assigned a nonresponse adjustment weight of  $w_i = \hat{p}_i^{-1}$ .

6. Next, we fit the measurement error model (1) to the data from responding individuals using method of moments. (For this measurement error model, method of moments provides similar estimates to those calculated using maximum likelihood.) We estimate both unweighted and weighted model parameters. The unweighted estimator of  $\mu_T$  is

$$\hat{\mu}_T^{unwt} = n_R^{-1} \sum_{i=1}^{n_R} \bar{X}_i,$$

where  $n_R$  is the number of responding individuals (those with  $R_i = 1$ ) and  $\bar{X}_i$  is the average of  $X_{i1}$  and  $X_{i2}$ . The weighted estimator of  $\mu_T$  is

$$\hat{\mu}_T^{wt} = \frac{\sum_{i=1}^{n_R} w_i \bar{X}_i}{\sum_{i=1}^{n_R} w_i},$$

where  $w_i$  is the nonresponse adjustment weight defined in Step 5. The unweighted and weighted estimators of  $\sigma_e^2$  are

$$\hat{\sigma}_{e,unwt}^2 = n_R^{-1} \sum_{i=1}^{n_R} \sum_{j=1}^2 (X_{ij} - \bar{X}_i)^2,$$

and

$$\hat{\sigma}_{e,wt}^2 = \frac{\sum_{i=1}^{n_R} \sum_{j=1}^2 w_i (X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^{n_R} w_i},$$

respectively. The unweighted and weighted estimators of  $\sigma_T^2$  are

$$\hat{\sigma}_{T,unwt}^2 = (n_R - 1)^{-1} \sum_{i=1}^{n_R} (\bar{X}_i - \bar{X}_{..})^2 - \frac{1}{2} \hat{\sigma}_{e,unwt}^2,$$

and

$$\hat{\sigma}_{T,wt}^2 = \frac{\sum_{i=1}^{n_R} w_i (\bar{X}_i - \bar{X}_{..})^2}{\left(\sum_{i=1}^{n_R} w_i\right) - 1} - \frac{1}{2} \hat{\sigma}_{e,wt}^2,$$

respectively, where  $\bar{X}_{..}$  is the overall mean of the  $X_{ij}$  for responding individuals.

7. We estimate  $\hat{\theta}_{naive}$  as

$$\hat{\theta}_{naive} = n_R^{-1} \sum_{i=1}^{n_R} I\{\bar{X}_i \geq 150\},$$

where  $I\{\bar{X}_i \geq 150\}$  is an indicator function that takes a value of 1 if  $\bar{X}_i$  is greater than or equal 150, and a value of 0 otherwise.

8. We estimate  $\hat{\theta}_{nr}$  as

$$\hat{\theta}_{nr} = \frac{\sum_{i=1}^{n_R} w_i I\{\bar{X}_i \geq 150\}}{\sum_{i=1}^{n_R} w_i},$$

where  $w_i$  is the nonresponse adjustment weight defined in Step 5.

9. To estimate  $\hat{\theta}_{me}$  we first generate 1,000 values of  $\tilde{X}_i$  from a normal distribution with mean  $\hat{\mu}_T^{unwt}$  and variance  $\hat{\sigma}_{T,unwt}^2$ . Then, the estimator is

$$\hat{\theta}_{me} = \frac{\sum_{i=1}^{1,000} I\{\tilde{X}_i \geq 150\}}{1,000}.$$

This approach, which adjusts for the measurement error in  $X_{ij}$ , is based on methods described in [Dodd et al. \(2006\)](#) and elsewhere.

10. To estimate  $\hat{\theta}_{full}$ , we use a similar approach to that described in Step 9, but use the weighted measurement error model parameter estimates instead of the unweighted estimates. We generate 1,000 values of  $\tilde{X}_i$  from a normal distribution with mean  $\hat{\mu}_T^{wt}$  and variance  $\hat{\sigma}_{T,wt}^2$ . Then, the estimator  $\hat{\theta}_{full}$  is

$$\hat{\theta}_{full} = \frac{\sum_{i=1}^{1,000} I\{\tilde{X}_i \geq 150\}}{1,000}.$$

This entire process (Steps 1–10) is repeated  $M$  times to account for simulation variation. For each of the  $M$  simulations and each of the four estimators, we calculate the bias as the difference between an estimate and the true value of the parameter  $\theta$ . The average bias across the  $M$  simulations is then calculated for each of the four estimators.

For the simulation, we set  $\mu_T = 4.20$  and  $\sigma_T^2 = 1.44$  which represent the mean and variance of time spent in MVPA during a typical week in the log scale. We use the log scale because physical activity data are often analyzed in the log scale ([Ferrari et al. 2007](#);

Tooze et al. 2013). For these choices of  $\mu_T$  and  $\sigma_T^2$ , the true value of  $\theta$  is about 25 percent and the ratio of  $\mu_T$  and  $\sigma_T^2$  is similar to the ratio found in other studies (Ferrari et al. 2007; Tooze et al. 2013).

We set  $\sigma_\varphi^2 = 0.04$  to estimate the response propensities (see Step 5 above). Values of  $\sigma_\varphi^2$  larger than 0.04 introduce increased uncertainty in the estimated response propensities which diminishes the effectiveness of the nonresponse adjustments. Although this may very well occur in practice, the goal of these simulations is to compare the effects of adjusting and not adjusting for measurement error and nonresponse when the adjustments are considered effective, without model misspecifications that could result in poorly estimated response propensities or measurement error model parameters.

For the remaining model parameters that subject the data to measurement error and nonresponse bias, we considered a range of scenarios. We let the measurement error variance,  $\sigma_e^2$ , range from 0 (no measurement error), to 8 (large measurement error) by 1-unit increments. With larger measurement error there will be more positive bias when estimating  $\theta$  without proper adjustments for the measurement error because there will be more variability in the data and consequently a higher percentage of cases above the threshold of 150 minutes of MVPA per week. We let the response propensity parameter,  $\alpha_1$ , range from  $-2$  to  $2$  by 0.5-unit increments. Negative values of  $\alpha_1$  will result in negative bias when estimating  $\theta$  (without accounting and adjusting for nonresponse) and positive values of  $\alpha_1$  will result in positive bias. The empirical values of  $\alpha_0$  are chosen to correspond with values of  $\alpha_1$ , so that each set of response propensity parameters gives an expected response rate of about 60 percent. In total, we considered 81 different simulation scenarios based on the nine values of  $\sigma_e^2$  and nine combinations of  $\alpha_0$  and  $\alpha_1$ .

Table A. National surveys with physical activity self-reports.

National survey	Estimated percentage of adults that adhere to physical activity guideline	Survey Year(s) in which estimate is based	Data collection mode	Link to questionnaire used to estimate MVPA	Source
BRFSS	65%	2007	Landline telephone	<a href="http://www.cdc.gov/brfss/questionnaires/pdf-ques/2007brfss.pdf">http://www.cdc.gov/brfss/questionnaires/pdf-ques/2007brfss.pdf</a>	Loustalot et al. (2009)
NHANES	51%	2003–2004	Field interview	<a href="https://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/sp_paq_c.pdf">https://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/sp_paq_c.pdf</a>	Troiano et al. (2008)
	60%	2005–2006		<a href="https://www.cdc.gov/nchs/data/nhanes/nhanes_05_06/sp_paq_d.pdf">https://www.cdc.gov/nchs/data/nhanes/nhanes_05_06/sp_paq_d.pdf</a>	Tucker et al. (2011)
NHIS	42%	1997–2004	Field interview	<a href="ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2004/english/QADULT.pdf">ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2004/english/QADULT.pdf</a>	Schoenborn and Stommel (2011)

Note: All physical activity measurements were self-administered. The NHIS estimated percentage of adults that adhere to physical activity guideline was averaged across the 1997–2004 survey years.

## 5. References

- Adams, S.A., C.E. Matthews, C.B. Ebbeling, C.G. Moore, J.E. Cunningham, J. Fulton, and J.R. Hebert. 2005. "The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity." *American Journal of Epidemiology* 161: 389–398. Doi: <https://doi.org/10.1093/aje/kwi054>.
- Ainsworth, B.E., W.L. Haskell, M.C. Whitt, M.L. Irwin, A.M. Swartz, S.J. Strath, W.L. O'Brien, D.R. Bassett, K.H. Schmitz, P.O. Emplaincourt, and D.R. Jacobs. 2000. "Compendium of Physical Activities: An Update of Activity Codes and MET Intensities." *Medicine and Science in Sports and Exercise* 32: S498–S504. Available at: <https://pdfs.semanticscholar.org/314e/dc8553c9a5920a14eb799b67c2a11e07b8bf.pdf> (accessed April 2017).
- Bassett, D.R., A.L. Cureton, and B.E. Ainsworth. 2000. "Measurement of Daily Walking Distance: Questionnaire versus Pedometer." *Medicine and Science in Sports and Exercise* 32: 1018–1023. Available at: <http://journals.lww.com/acsm-mssse/pages/articleviewer.aspx?year=2000&issue=05000&article=00021&type=abstract> (accessed April 2017).
- Beyler, N., S. James-Burdumy, M. Bleeker, J. Fortson, and M. Benjamin. 2015. "Estimated Distributions of Usual Physical Activity During Recess". *Medicine and Science in Sports and Exercise* 46: 1197–1203. Doi: <https://doi.org/10.1249/MSS.0000000000000535>.
- Carroll, R.J., D. Ruppert, and L.A. Stefanski. 2006. *Measurement Error in Nonlinear Models*. New York: CRC Press.
- Crouter, S.E., K.G. Clowers, and D.R. Bassett. 2006. "A Novel Method for Using Accelerometer Data to Predict Energy Expenditure." *Journal of Applied Physiology* 100: 1324–1331. Doi: <http://dx.doi.org/10.1152/jappphysiol.00818.2005>.
- Dodd, K.W., P.M. Guenther, L.S. Freedman, A.F. Subar, V. Kipnis, D. Midthune, J.A. Tooze, and S.M. Krebs-Smith. 2006. "Statistical Methods for Estimating Usual Intake of Nutrients and Foods: A Review of the Theory." *Journal of the American Dietetic Association* 106: 1640–1650. Doi: <http://doi.org/10.1016/j.jada.2006.07.011>.
- Ferrari, P., C. Friedenreich, and C.E. Matthews. 2007. "The Role of Measurement Error in Estimating Levels of Physical Activity." *American Journal of Epidemiology* 166: 832–840. Doi: <https://doi.org/10.1093/aje/kwm148>.
- Groves, R.M. and L. Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74: 849–879. Doi: <https://doi.org/10.1093/poq/nfq065>.
- Hill, A., J. Roberts, P. Ewings, and D. Gunnell. 1997. "Non-Response Bias in a Lifestyle Survey." *Journal of Public Health* 19: 203–207. Doi: <https://doi.org/10.1093/oxford-journals.pubmed.a024610>.
- Lessler, J.T. and W.D. Kalsbeek. 1992. *Nonsampling Error in Surveys*. New York: Wiley.
- Loustalot, F., S.A. Carlson, J.E. Fulton, J. Kruger, D.A. Galuska, and F. Lobelo. 2009. "Prevalence of Self-Reported Aerobic Physical Activity among US States and Territories—Behavioral Risk Factor Surveillance System, 2007." *Journal of Physical Activity and Health* 6: S9–S17. Doi: <http://dx.doi.org/10.1123/jpah.6.s1.s9>.

- Matthews, C.E. 2002. "Use of Self-Report Instruments to Assess Physical Activity." In *Physical Activity Assessments for Health Related Research*, edited by G.J. Welk, 107–123. Champaign: Human Kinetics.
- Matthews, C.E. 2005. "Calibration of Accelerometer Output for Adults." *Medicine and Science in Sports and Exercise* 37: S512–S522. Doi: <http://dx.doi.org/10.1249/01.mss.0000185659.11982.3d>.
- Nusser, S.M., N.K. Beyler, G.J. Welk, A.L. Carriquiry, W.A. Fuller, and B. King. 2012. "Modeling Errors in Physical Activity Recall Data." *Journal of Physical Activity and Health* 9: S56–S67. Doi: <http://dx.doi.org/10.1123/jpah.9.s1.s56>.
- Sallis, J.F. and B.E. Saelens. 2000. "Assessment of Physical Activity by Self-Report: Status, Limitations and Future Directions." *Research Quarterly for Exercise and Sport* 71: 1–14. Doi: <http://dx.doi.org/10.1080/02701367.2000.11082780>.
- Schoenborn, C.A. and M. Stommel. 2011. "Adherence to the 2008 Adult Physical Activity Guidelines and Mortality Risk." *American Journal of Preventive Medicine* 40: 514–521. Doi: <http://doi.org/10.1016/j.amepre.2010.12.029>.
- Smith, C. and D. Nutbeam. 1990. "Assessing Non-Response Bias: A Case Study from the 1985 Welsh Heart Health Survey." *Health Education Research* 5: 381–386. Doi: <https://doi.org/10.1093/her/5.3.381>.
- Tooze, J.A., R.P. Troiano, R.J. Carroll, A.J. Moshfegh, and L.S. Freedman. 2013. "A Measurement Error Model for Physical Activity Level as Measured by a Questionnaire with Application to the 1999–2006 NHANES Questionnaire." *American Journal of Epidemiology* 177: 1199–1208. Doi: <http://dx.doi.org/10.1093/aje/kws379>.
- Troiano, R.P., D. Berrigan, K.W. Dodd, L.C. Masse, T. Tilert, and M. McDowell. 2008. "Physical Activity in the United States Measured by Accelerometer." *Medicine and Science in Sports and Exercise* 40: 181–188. Doi: <http://dx.doi.org/10.1249/mss.0b013e31815a51b3>.
- Tucker, J.M., G.J. Welk, and N.K. Beyler. 2011. "Physical Activity in U.S. Adults: Compliance with the Physical Activity Guidelines for Americans." *American Journal of Preventive Medicine* 40: 454–461. Doi: <http://dx.doi.org/10.1016/j.amepre.2010.12.016>.
- U.S. Department of Health and Human Services. 2008. *2008 Physical Activity Guidelines for Americans*. Atlanta: HHS. Available at: <http://health.gov/paguidelines> (accessed November 2015).
- Van Loon, A.J.M., M. Tijhuis, H.S.J. Picavet, P.G. Surtees, and J. Ormel. 2003. "Survey Non-Response in the Netherlands: Effects on Prevalence Estimates and Associations." *Annals of Epidemiology* 13: 105–110. Doi: [http://doi.org/10.1016/S1047-2797\(02\)00257-0](http://doi.org/10.1016/S1047-2797(02)00257-0).
- Vink, J.M., G. Willemsen, J.H. Stubbe, C.M. Middeldorp, R.S. Ligthart, K.D. Baas, H.J. Dirkszager, E.J. de Geus, and D.I. Boomsma. 2004. "Estimating Non-Response Bias in Family Studies: Application to Mental Health and Lifestyle." *European Journal of Epidemiology* 19: 623–630. Doi: <http://dx.doi.org/10.1023/B:EJEp.0000036814.56108.66>.
- Warnecke, R.B., T.P. Johnson, N. Chavez, S. Sudman, D.P. O’rourke, L. Lacey, and J. Horm. 1997. "Improving Question Wording in Surveys of Culturally Diverse

Populations.” *Annals of Epidemiology* 7: 334–342. Doi: [https://doi.org/10.1016/S1047-2797\(97\)00030-6](https://doi.org/10.1016/S1047-2797(97)00030-6).

Welk, G.J. 2002. “Introduction to Physical Activity Research.” In *Physical Activity Assessments for Health Related Research*, edited by G.J. Welk, 3–18. Champaign: Human Kinetics.

Received January 2016

Revised March 2017

Accepted April 2017