MINNESOTA POPULATION CENTER

UNIVERSITY OF MINNESOTA

Implications of Measurement: Comparing ATUS Estimates of Physical Activity to NHANES

Rachelle Hill†
U.S. Census Bureau

Sarah Flood
University of Minnesota

Kari Williams
University of Minnesota

March 2020

Working Paper No. 2020-01
DOI: https://doi.org/10.18128/MPC2020-01

**Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau or the University of Minnesota. All results have been reviewed to ensure that no confidential information is disclosed.

†Address correspondence to Rachelle Hill (email: rachelle.hill@census.gov). Support for this work was provided by the Minnesota Population Center at the University of Minnesota (P2CHD041023) and the IPUMS Health Surveys Project (R01HD046697).

Abstract

Measuring physical activity reliably and accurately is challenging. The gold standard is limited because it is resource intensive. Accelerometer data has been used as a more accessible option but other less explored data such as time diary data could be a good alternative. Understanding how estimates differ by measurement strategy, including time diary data and accelerometer data, is thus an important contribution to the limited research base about physical activity measurement. Using nationally representative data from the 2003-2006 National Health and Nutrition Examination Survey (NHANES) and the 2006-2008 American Time Use Survey (ATUS), we investigate the following questions. First, how do recall estimates of time spent in physical activity in the ATUS compare to NHANES estimates based on accelerometer data? And, to what extent do these estimates vary by demographic characteristics?

The literature investigating patterns and prevalence of physical activity is expansive. This is partially a reflection of consensus regarding the role of physical activity in reducing morbidity and mortality. Furthermore, given the centrality of physical activity for improving and maintaining individual health and well-being, physical activity has been measured in a variety of ways. Less well understood, however, is the extent to which different ways of measuring physical activity yield similar estimates and correlates of physical activity.

We compare two approaches to measuring physical activity—via accelerometer and time diary—both of which have their merits. Accelerometer data are an objective means of measuring physical activity; data collection requires minimal self-reflection and evaluation by individual participants who wear a device that tracks their movement. In contrast, time diary data are based on individual reports of the activities they engage in over, for example, a 24-hour period. While time diary data are less commonly used in physical activity research and not collected specifically to measure physical activity, the methodology is less likely to influence behavior or reporting of it compared to direct questions about physical activity. Both types of data have limitations. Accelerometer data collection is burdensome for a large sample of participants because of effort required to manage the devices and ensuring participants wear the device once they've been selected, in addition to the limited contextual detail. Time diary data provide limited visibility into long-term physical activity patterns and intensity given the typical one-day diary design. Nonetheless, we argue that time diary data are a key, underutilized resource for research on physical activity.

Using nationally representative data from the 2003-2006 National Health and Nutrition Examination Survey (NHANES) and the 2006-2008 American Time Use Survey (ATUS), we investigate the following questions. First, how do time diary estimates of time spent in physical activity from the ATUS compare to NHANES estimates based on accelerometer data? And to what extent do the differences in these estimates vary by demographic characteristics?

Background

The duration, intensity, and frequency of physical activity are often discussed in scholarly and popular media as important indicators of current and future physical health, in part because exercise has been linked to decreased risk of mortality, heart disease, hypertension, colon and breast cancer, diabetes, and depression (U. S. Department of Health and Human Services 1996; U.S. Department of Health and Human Services 2010). Despite well-documented benefits, time spent in physical activity is low (especially when limited to bouts of 10 minutes or more), peaking in childhood and quickly declining across the life course for both men and women as well as for all racial and ethnic groups (Troiano et al. 2008). Few adults meet the minimum recommendations for physical activity (Tucker, Welk, and Beyler 2011).

Physical activity is measured using several methodologies, including the doubly labeled water method, questionnaires, accelerometers, and time diaries. The gold standard for physical activity measurement is the doubly labeled water method (Westerterp 2017). The doubly labeled water method measures energy expenditure by having participants ingest water containing uncommon isotopes and comparing their elimination rates over time. This method is costly and logistically challenging for larger samples and requires additional data collection methods to gain insight into the context surrounding physical activity choices.

Accelerometer data are an alternative method of objectively measuring physical activity (Prince et al. 2008). Accelerometers, in particular those used by NHANES participants, are small devices worn on the hip of a study participant that track the duration and intensity of common activities used for locomotion (e.g., walking or running). Though collecting physical activity data via accelerometer is more cost effective and includes more contextual information than the doubly labeled water method, the resources necessary to distribute, maintain, and collect devices are a barrier to collecting accelerometer data among large, population-representative samples (Prince et al. 2008). Additionally, accelerometers require participants to cooperate by consistently wearing a device during waking hours and without additional survey information, little is known about the activities engaged in or for what purpose. For example, when the device collects no information, presumably because it is not being worn, there is no way to determine why the participant is not wearing the device, the level of activity she/he was engaged in for these unmeasured periods, or the impact on non-response bias. Such challenges make other measurement approaches to physical activity highly attractive. Alternatives might include self-report measures via stylized questions where respondents report how frequently they engage in a given physical activity or via time diaries where respondents report what they do over the course of the day.

Self-report data rely on the recollections of respondents about their physical activity. Self-reports, usually collected through interviews or questionnaires, are typically low cost and low burden, which can facilitate larger and more representative samples. Recall and response bias represent potential challenges to reliable data as participants may over- or under-report their physical activity because of social desirability, inaccurate memories, or lack of clarity about which physical activities are to be included (Adams et al. 2005; Prince et al. 2008). Most research that compares self-reports of physical activity to more objective measures use stylized summary questions that assess the total time spent in all types of activities during a set period of time.[1] Such prior research has demonstrated both over- and underestimates of time in physical activity that vary by demographics when using stylized measures and accelerometer data (Adamo et al. 2009; Kowalski et al. 2012; Prince et al. 2008; Sabia et al. 2014; Tucker et al. 2011), but rarely does this comparative research assess self-reported time diary data.

Time diaries are a type of self-report data and an alternative to stylized summary questions. Time diary surveys ask individuals to describe all of the activities they participated in over a given period of time. A strength of the time diary method is that individuals are not asked about a specific type of activity. Due to the breadth of activities participated in during a given day, individuals are less likely to be biased towards a specific activity (Juster and Stafford 1985). Though knowledge about how time diary estimates compare to accelerometer estimates of physical activity for the broad US population is scant, similar 24-hour recall studies that also use accelerometer have shown the potential for accurate and reliable estimates of physical activity (Matthews et al. 2018). The previous research that has

[1] Previous versions of this research included physical activity duration based on stylized questions from the National Health Interview Survey (IPUMS NHIS [Blewett, Drew, King and Williams, 2019]) as a point of comparison. However, the NHIS stylized questions about physical activity do not inquire about exercise bouts of less than 10 minutes. Additionally, NHIS allows respondents to self-select the reference period for their physical activity (i.e., bouts per day, week, month, or year), requiring decisions about how to allocate reported physical activity to a single day. Given these differences and the relatively robust literature comparing stylized questions to accelerometer data, we eliminated NHIS data from the analyses presented here.

investigated time diary type data and accelerometer data has done so in small samples of children or the elderly (Bringolf-Isler et al. 2012; Jago et al. 2005; Rutgers, Klijn, and Deurenberg 1997). Gershuny and colleagues (2017) draw on a small sample of adults from the UK to compare accelerometer data to time diary and find initial evidence that time diary data may be a strong alternative. While the extant literature leads us to believe that the objective accelerometer and subjective time diary data will have notable similarities, it is unknown if there are systematic differences between time diary and accelerometer data for the broader US population.

We extend this research by comparing accelerometer and time diary approaches to measuring physical activity in NHANES and ATUS. We answer the following research questions: How do recall estimates of time spent in physical activity in the ATUS compare to NHANES estimates based on accelerometer data? And to what extent do demographic and health characteristics explain differences in physical activity estimates?

Data and Methods

We use data from NHANES and ATUS to perform our comparisons of physical activity. Our key focus is on understanding the extent to which ATUS time-diary based estimates of physical activity are similar or different from NHANES accelerometer based estimates, including how close estimates are after controlling for demographic and health differences in sample composition. For comparability across datasets and due to our focus on working-aged adults, we limit our analytic sample to respondents aged 18 to 64 with no missing data on physical activity or other controls included in our models.

NHANES

We use data from the 2003 to 2006 National Health and Nutrition Examination Survey (NHANES) (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2013). NHANES is a study of health status, health behaviors, and nutrition among a nationally representative sample of the non-institutionalized civilian population in the United States conducted by the National Center for Health Statistics (NCHS). Data were collected via household interviews and physical examinations completed in mobile examination centers. The survey oversamples blacks and Mexican Americans, low-income individuals, and individuals within the same household. Key for our purpose are the accelerometer data, which captures “the magnitude of acceleration” for each minute the small device is worn on the hip (Centers for Disease Control and Prevention 2007). These intensity values are then converted to metabolic equivalents (MET). Following their participation in the survey, NHANES participants are asked to schedule the examination. During the examination participants are invited to wear an accelerometer for seven days during waking hours.[2] The full 2003 to 2006 NHANES sample included 20,470 respondents. Our analytic sample is 3,679 individuals ages 18-64, who wore the physical activity monitor for at least 10 hours on four days and contain full information on our independent variables.[3] We randomly choose a single day of accelerometer data for each respondent to more accurately compare to the data from ATUS respondents.

[2] Respondents are asked not to wear the accelerometer while swimming and sleeping. Also, accelerometers like the one used here, are not designed to effectively capture effort exerted while biking.

[3] Of the interviewed sample, 95.7 percent of the sample is included in the examined sample. Of the examined sample, 85 percent reported at least four days of data from the accelerometer. The remaining respondents are

ATUS

We use integrated American Time Use Survey (ATUS) data for our analyses (Hofferth, Flood, and Sobek 2015). The ATUS is a time diary study of a nationally representative sample of Americans and has been collected annually since 2003. ATUS sample members are invited to complete the survey following their exit from the Current Population Survey (CPS), which is a household survey of the civilian, non-institutionalized population. One individual aged 15 or older per former CPS participating household is randomly selected to respond to the ATUS during the two to five months following their exit from the CPS. ATUS respondents report the activities they engaged in over a 24-hour period from 4:00 a.m. of yesterday until 3:59 a.m. of the reporting day (diary day), as well as where, when, and with whom activities are done. Over 400 detailed activity codes are represented in the three-tier six-digit activity coding scheme. Data are collected on all days of the week, and weekends are oversampled. Hispanic and Black respondents as well as those living in households with children are also oversampled. Though the data may not typify respondents' daily activities, aggregations of the data are representative of the American population. We restricted our ATUS data to 2006-2008 to increase comparability with NHANES while also including self-reported health and BMI. The analytic sample included 37,914 respondents, 25,384 of whom are 18-64 and contain full information on our independent variables.

Measures

Physical Activity Measures

As highlighted earlier, NHANES and ATUS assess physical activity in different ways, where NHANES uses an accelerometer (or physical activity monitor), ATUS uses a time diary. Each method is described below.

NHANES respondents wear the accelerometer (Actigraph), which is calibrated to measure physical movement and is meant to objectively measure participants' physical activity levels (Centers for Disease Control and Prevention 2005). Those younger than six and those subjects in wheelchairs or with other physical impairments that prevent walking are excluded from this portion of the study. Participants were asked to wear the accelerometer on their right hip during their waking hours for seven days; the accelerometer captured intensity of movement at 1-minute intervals. The raw data from NHANES were first edited using programs available by the Applied Research Cancer and Population Sciences division of the National Cancer Institute (2013) to make the physical activity estimates comparable to other datasets as well as to reweight the accelerometer data to match the larger NHANES examined sample. For each individual we randomly choose one of the days the accelerometer was worn to increase comparability to the ATUS.[4] Each minute of physical activity in the NHANES has a metabolic equivalent greater than three (Troiano 2008).

ATUS respondents report a time diary about what activities they did in the reference period as well as detailed information about each activity episode from which we derive our physical activity measure. Each episode of activity is recorded separately and the minutes spent in that activity are recorded. Physical activity in the ATUS includes all time the respondent was exercising or playing sports (hunting

---

excluded from our analysis because of missing data on the control variables or they were excluded based on sample criteria.

[4] The randomly chosen days are even distributed across all days of the week with percentages varying between 14.09 percent and 14.7 percent for each day.

and fishing are excluded) *or* walking as a mode of transportation. We exclude episodes of swimming and biking from the ATUS measure to increase comparability with NHANES, as the accelerometer does not capture biking and swimming. Importantly, because the time diary only covers one 24-hour period, some respondents will not report any physical activity—either exercising or walking somewhere—because exercise may not be a daily activity for that individual. Also, though some jobs may be physically demanding, we do not have detail on physical activity during the workday.

We calculated two measures of daily physical activity: total sum minutes of physical activity (TSM) and total conditional minutes of physical activity (TCM). Both are continuous measures of the total time spent in physical activity per day as defined for each of the data collection methods. TSM includes all minutes spent in physical activity on the preselected day while TCM is limited to episodes of physical activity that were 10 or more minutes[5] in duration (i.e., time included in this measure of physical activity is conditional on the episode being 10 minutes or more). We refer to these episodes as bouts similar to previous literature.

Demographic Characteristics

We include controls for gender, age, race, native- or foreign-born status, marital status, education, income, employment status, self-reported health and body mass index (BMI). *Gender* is coded as male (reference) and female. *Age* of the respondent is measured as a categorical variable: 19-24 (reference), 25-34, 35-44, 45-54, and 55-64. *Race* is coded as four dichotomous variables: White, non-Hispanic (reference); Black, non-Hispanic; Other, non-Hispanic; and Hispanic. *Immigrant* indicates whether the respondent was native (reference) or foreign born. *Marital status* differentiates between married (reference), cohabiting, widowed/divorced/separated, and never married. *Education* is coded into four dichotomous variables: less than high school (reference), high school degree, some college, and college degree or more. *Income* is coded into three broad categories: up to $34,999 (reference), $35,000-$74,999, and $75,000 and higher. *Employment status* differentiates into full time (reference), part time (1-34 hours per week), non-employed, and self-employed. *Self-reported health* distinguishes between excellent or very good, good (reference), and fair or poor. *BMI* includes categories for underweight (<18.5), normal weight (18.5-24.9), overweight (25-29.9), and obese (30+); normal weight is the reference.

Analytic Method

Our analytic approach is multi-step and examines the sample composition, estimates the unadjusted averages of time spent in physical activity across the population for each dataset, and predicts the adjusted physical activity estimates that account for demographic differences across surveys. We first compare demographic characteristics across the weighted samples. We then pool our analytic samples and use OLS regression to predict TSM and TCM, controlling for demographic characteristics. Finally, we include interaction terms between each of our independent variables and ATUS to identify where the datasets, and by proxy the measurement methods, are differentially related to the physical activity estimates produced by the OLS models. We complete the analysis by conducting Wald Tests of significance for the interaction effects. Wald tests are an approach to testing the statistical significance

[5] The program provided by the Division of Cancer Control and Population Sciences of the National Cancer Institute (2013) allows for the readings from the accelerometer to dip below the moderate or vigorous activity threshold for up to two minutes before it determines that an episode is complete.

of variables or groups of variables and therefore can more easily determine if model fit improves if an interaction term with multiple categories is included in the model. We apply weights normalized to the original samples in all analyses.

Results

Descriptive Statistics

In Table 1 we compare the sample composition of NHANES and ATUS. Using the sample weights, each sample is meant to be nationally representative and should have similar sample characteristics. Few demographic characteristics vary by survey. Approximately 50 percent of the sample is female in NHANES and ATUS. There are also similar proportions of each race/ethnicity, immigrant status. There is less variation than expected in BMI categories (e.g. 33.2 percent and 32.5 percent are normal weight in NHANES and ATUS, respectively) especially considering that weight and height are self-reported measures in ATUS, but are measured objectively during the exam in NHANES. There is more difference across the datasets than expected in marital status, employment status, education, and income. The proportion of respondents who report being in excellent health is higher in ATUS (55.1 percent) compared to NHANES (49.3 percent) respondents.

Finally, we see differences in minutes spent in average daily physical activity. The average total sum minutes (TSM) spent in physical activity is 28.8 minutes for NHANES and 15.5 minutes for ATUS. In contrast, the total conditional minutes (TCM) spent in physical activity—based on a minimum threshold of 8 minutes of physical activity during a ten-minute span—is 9.2 for NHANES and 14.9 minutes per day for ATUS. Notably, the ATUS estimates are very similar regardless of whether or not the physical activity includes all time or only conditional bouts, while the NHANES estimate is larger when all time is included as compared to only conditional bouts. The large differences between TSM and TCM in NHANES, as compared to ATUS may reflect how individuals report their time. Rarely do individuals report time in 1-minute intervals. Instead, people commonly heap their responses on 5 and 10 minute increments and tend to omit 1-4 minute episodes from their diaries.

Table 1 about here.

OLS Regression Models

The first set of columns in Table 2 show the base OLS model estimates for TSM and TCM using the pooled NHANES and ATUS data. The associations between demographic characteristics and physical activity are consistent with the literature. We generally find that older individuals, women, and those with lower education and higher BMI spend less time in physical activity on average. Our key focus, however, is on differences between NHANES and ATUS. The ATUS indicator in the model shows that estimated minutes spent in physical activity replicate the expected differences between measurement methods seen in the descriptive statistics, even net of sample characteristics. The negative coefficient on the ATUS indicator for TSM reflects higher NHANES estimates for total minutes, and the positive coefficient for of the ATUS dataset for the TCM measure reflects higher estimates from ATUS compared to NHANES when only including bouts of 10 minutes or more. Table 2 includes the Wald tests for each set of coefficients. The results of the Wald tests, which test the statistical significance of the combined effect of the categories within a variable, show that race, marital status, education, BMI, and ATUS are statistically significant predictors of both TSM and TCM. Gender is not a statistically significant predictor of TCM while family income is.

Table 2 about here.

Interaction between independent variables and datasets

Columns three and four in Table 2 show the interacted models that include interactions between the ATUS indicator and each of the independent variables. There are significant differences across datasets in the effect of demographic and health characteristics on TSM and TCM. Significant interactions between the ATUS indicator and the demographic and health characteristics indicate that the dataset (a proxy for the measurement method used) differentially impacts the estimates for persons with these demographic characteristics, on average. The relationships between several demographic characteristics, including marital status, employment status, and occupational activity and physical activity, are statistically different in ATUS versus NHANES. These differences hold for both TSM and TCM. Some associations with physical activity are statistically different when looking at TSM (e.g. age, race, weekend days, and BMI interactions) while other relationships differ when looking at TCM (e.g. gender, income, and self-reported health).

Figure 1 illustrates the statistically significant relationships between the dataset, demographic characteristic, and TCM. Here we focus on gender, marital status, family income, and self-reported health given the differential treatment of work in the two datasets. We also focus on TCM because the baseline differences between the datasets is smaller. As the figure makes clear, the differences by demographic categories are greater in the ATUS data than the NHANES data. In addition, the increasing TCM for income and decreasing TCM for self-rated health with the ATUS data are not seen in the NHANES data.

Figure 1 about here

Conclusion

Physical activity is an important indicator of current and future individual health. Scholars have invested a great deal of time and energy in understanding who engages in physical activity as well as the outcomes associated with different levels of physical activity. Despite its importance for understanding health, physical activity is measured in many ways. To further our understanding of the differences associated with physical activity measurement, we compare two different techniques for measuring physical activity using nationally representative samples. A comparison of descriptive statistics across NHANES and ATUS illustrates similarities and differences in the samples. The samples are in many ways very similar including gender, age, and racial/ethnic composition. However, notable differences exist in the sample composition of the datasets including marital status, income, education, employment status, and self-reported health. Such differences are important for understanding subsequent results. In line with the different measurement approaches, we also find that time spent in physical activity differs by dataset though not in the ways we might have expected. After controlling for demographic characteristics, the differences between NHANES and ATUS are minimally changed, in particular when focusing on TCM. It may be unsurprising that TSM differences are greater when one considers the ability of the accelerometer to capture a minute of rushing to the bus or after a child, which has a greater chance of going unreported in the time diary data. TCM may map more closely onto how individual respondents report their diary day.

Overall we see that controlling for demographic characteristics reinforces a broader pattern that differences in time spent in physical activity do vary by measurement type. Despite the similar estimates of TCM, the significant interaction between dataset and many of our covariates highlights that the accelerometer versus time diary measurement methods are differentially related to physical activity estimation for specific groups. Specifically, the differences between the estimated time spent in physical activity by gender, marital status, income, employment status, occupational activity, and self-reported health are statistically significant in the ATUS but not in the NHANES. Such differences deserve a closer look in future research.

The differences in measurement of physical activity during work time are one possible reason why employment status, occupational activity, and even weekend effects are different across the datasets. ATUS does not ask for additional details regarding how physically active individuals are during work hours; as such, that time is not included as physical activity. However, physical activity during work hours may be captured by the accelerometer in NHANES. In addition, the differences in ATUS and NHANES samples by self-reported health, marital status, and, potentially, age may also help explain the different relationships between these variables in the NHANES versus the ATUS.

Digging deeper into the potential effects of limiting physical activity assessment to non-work hours, we focus on gender, marital status, family income, and self-reported health. We find that patterns in TCM by income and self-rated health in the ATUS data, do not occur in the NHANES data. This difference raises interesting questions about health status, economic advantage, and physical activity. Further exploration will need to consider why these characteristics are associated with different patterns across the datasets and how much a role occupation may play in these patterns.

Differences by gender, marital status, family income and self-reported health highlight several key points as well as areas for future research. While accelerometer data offer a direct measure of physical activity and are, therefore, objective and valuable, accelerometer measurement is not without drawbacks. The cost and interpretability of data are challenges that may prevent broader use of accelerometers for physical activity measurement and data collection. Participant burden and participant compliance are also potentially problematic. Researchers must make choices about what length of bout should be considered in measuring physical activity, how many valid days of wear are necessary to create a representative week of physical activity for respondents, and what type and intensity of physical activity might be happening while persons are not wearing the devices. This coupled with sample differences in self-reported health and other demographic characteristics, as well as the oversampling within households that we are unable to control for raise questions about typical statistical assumptions that are made when using both of these datasets like missing at random and adequately adjusting for multistage sampling.[6] Though previous research has examined nonresponse in the NHANES by comparing the examined versus the NHANES examination non-responders and found few differences (Forthofer 1983), it's unclear if similar results would be found when limiting the analysis to accelerometer participants and non-responders. Future research may explore how non-compliers

[6] Though weights are applied to the data for analysis, survey documentation state that oversampling within households is not explicitly adjusted for in the weights. Instead, weights adjust for age and gender which differs across household respondents (National Center for Health Statistics 2012).

differ from those persons who wear the device as directed, or if days when the device is worn differ significantly from non-wear days.

A comparison of ATUS to NHANES estimates indicates that persons are likely not over-reporting physical activity when asked about episodic activities in a time diary format. While social desirability response bias or forgetfulness may affect results, the structural nature of the ATUS limits bias towards certain activities compared to surveys that ask about the time typically allocated to a specific type of activity. However, because ATUS only provides a one-day snapshot for each respondent and may not typify her/his normal activities, it is important to consider ATUS estimates in the aggregate.

The similar patterns regardless of data source indicate that while the specific predicted minutes spent in physical activity differ across different data collection methods, there is value in each of these datasets and merit to each of the measurement methods depending on the research objective. Research with a greater focus on metabolic output or strong ties to other indicators of health should use the NHANES or other accelerometer data; this might also include research exploring the effect of minutes spent in physical activity on diabetes or variation in metabolic output across the day. ATUS is best used for population activity and can provide valuable insight into broader activity habits in addition to temporal and social context without priming respondents towards physical activity. ATUS is also better suited to study subpopulations that may not be well enough represented for subpopulation analysis in a smaller dataset like NHANES. Regardless of dataset used, accounting for demographics such as gender, age, race, marital status, employment status, income, occupational activity, BMI, and self-reported health are essential to account for differences across the population and datasets.

References

Adamo, Kristi B., Stéphanie A. Prince, Andrea C. Tricco, Sarah Connor-Gorber, and Mark Tremblay. 2009. "A Comparison of Indirect versus Direct Measures for Assessing Physical Activity in the Pediatric Population: A Systematic Review." *International Journal of Pediatric Obesity* 4(1):2–27.

Adams, Swann Arp et al. 2005. "The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity." *American Journal of Epidemiology* 161(4):389–98.

Applied Research Cancer Control Population Sciences. 2013. "SAS Programs for Analyzing NHANES 2003-2004 Accelerometer Data." Retrieved March 1, 2013 (http://appliedresearch.cancer.gov/nhanes_pam/).

Blewett, Lynn A., Julia A. Rivera Drew, Miriam L. King and Kari C.W. Williams. IPUMS Health Surveys: National Health Interview Survey, Version 6.4 [dataset]. Minneapolis, MN: IPUMS, 2019. https://doi.org/10.18128/D070.V6.4

Bringolf-Isler, Bettina et al. 2012. "Measuring and Validating Physical Activity and Sedentary Behavior Comparing a Parental Questionnaire to Accelermeter Data and Diaries." *Pediatric Exercise Science* 24(2):229–45.

Centers for Disease Control and Prevention. 2007. "NHANES 2005-2006: Physical Activity Monitor Data Documentation, Codebook, and Frequencies." Retrieved July 20, 2018 (https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/PAXRAW_D.htm).

Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS). 2013. *National Health and Nutrition Examination Survey Data*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Forthofer, Ronald N. 1983. "INVESTIGATION OF NONRESPONSE BIAS IN NHANES II." *American Journal of Epidemiology* 117(4):507–15.

Gershuny, Jonathan et al. 2017. "CAPTURE24: Testing Self-Report Time-Use Diaries against Objective Instruments in Real Time."

Hofferth, Sandra L. 2006. "Response Bias in a Popular Indicator of Reading to Children." *Sociological Methodology* 36(1):301–15.

Hofferth, Sandra L., Sarah M. Flood, and Matthew Sobek. 2015. *American Time Use Survey Data Extract System: Version 2.5 [Dataset]*. College Park, MD: University of Maryland and Minneapolis, MN: University of Minnesota.

Jago, Russell, Cheryl B. Anderson, Tom Baranowski, and Kathy Watson. 2005. "Adolescent Patterns of Physical Activity: Differences by Gender, Day, and Time of Day." *American Journal of Preventive Medicine* 28(5):447–52.

Juster, F. Thomas and Frank P. Stafford. 1985. *Time, Goods, and Well-Being*. Ann Arbor, MI: Institute for Social Research.

Kowalski, Kristina, Ryan Rhodes, Patti-Jean Naylor, Holly Tuokko, and Stuart MacDonald. 2012. "Direct and Indirect Measurement of Physical Activity in Older Adults: A Systematic Review of the Literature." *International Journal of Behavioral Nutrition and Physical Activity* 9(1):148.

Matthews, Charles E. et al. 2018. "Measurement of Active and Sedentary Behavior in Context of Large Epidemiologic Studies." *Medicine and Science in Sports and Exercise* 50(2):266–76.

National Center for Health Statistics, ed. 2012. *National Health and Nutrition Examination Survey: Sample Design,1999-2006*. Hyattsville, Md: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Prince, Stéphanie A. et al. 2008. "A Comparison of Direct Versus Self-Report Measures for Assessing Physical Activity in Adults: A Systematic Review." *International Journal of Behavioral Nutrition and Physical Activity* 5(1):56.

Robinson, John P. and Ann Bostrom. 1994. "The Overestimated Workweek? What Time Diary Measures Suggest." *Monthly Labor Review* 117(8):11–23.

Rutgers, Carola J., Marja J. C. Klijn, and Paul Deurenberg. 1997. "The Assessment of 24-Hour Energy Expenditure in Elderly Women by Minute-by-Minute Heart Rate Monitoring." *Annals of Nutrition and Metabolism* 41(2):83–88.

Sabia, Séverine et al. 2014. "Association Between Questionnaire- and Accelerometer-Assessed Physical Activity: The Role of Sociodemographic Factors." *American Journal of Epidemiology* 179(6):781–90.

Troiano, Richard P. et al. 2008. "Physical Activity in the United States Measured by Accelerometer." *Medicine & Science in Sports & Exercise* 40(1):181–88.

Tucker, Jared M., Gregory J. Welk, and Nicholas K. Beyler. 2011. "Physical Activity in U.S. Adults: Compliance with the Physical Activity Guidelines for Americans." *American Journal of Preventive Medicine* 40(4):454–61.

U. S. Department of Health and Human Services. 1996. *Physical Activity and Health: A Report of the Surgeon General*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

U.S. Department of Health and Human Services. 2010. "Healthy People 2020 - Improving the Health of Americans." *Healthy People 2020*. Retrieved December 29, 2010 (http://www.healthypeople.gov/2020/default.aspx).

Westerterp, Klaas R. 2017. "Doubly Labelled Water Assessment of Energy Expenditure: Principle, Practice, and Promise." *European Journal of Applied Physiology* 117(7):1277–85.

Table 1. Comparison of Demographic Characteristics by Survey

	NHANES 2003-2006	ATUS 2006-2008	
Daily Physical Activity (mean minutes)	28.80	15.47	1
Daily Physical Activity (conditional on 8 minutes in sequence)	9.20	14.88	1
BMI			
Under weight (<=18.5)	1.75	1.62	
Normal weight (18.5-24.9)	33.22	32.49	
Overweight (25-29.9)	32.65	33.65	
Obese (30 or greater)	32.38	32.24	
Self-Reported Health			
Very Good/Excellent	49.27	55.08	1
Good	35.63	30.1	1
Fair/Poor	11.74	13.99	1
Gender			
Women	50.96	50.58	
Men	49.04	49.42	
Age			
19-24	11.12	12.96	1
25-34	22.9	21.86	
35-44	23.86	23.31	
45-54	26.24	23.97	
55-64	15.87	17.9	1
Race			
Non-Hispanic White	70.19	68.17	
Non-Hispanic Black	11.93	11.57	
Hispanic	12.13	14.63	
Other	5.75	5.63	
Immigrant			
Born outside of the U.S.	15.88	16.08	
Born in the U.S	84.12	83.92	
Marital Status			
Married	60.27	59.06	
Cohabiting	7.98	4.52	1
Widowed/Divorced/Separated	13.16	13.06	
Never Married	18.55	23.36	1
Educational Attainment			
Less than High School Degree	12.72	11.4	
High School Degree or Equivalent	24.31	30.34	1
Some College or Associate's Degree	34.29	28.57	1
College Degree or Higher	28.68	29.69	
Family Income			
Up to $34,999	33.05	29.79	1
$35,000 to $74,999	36.56	36.24	
$75,000 and up	30.39	33.97	
Employment status			
Part-time (less than 35 hours per week)	7.51	11.38	1
Full-time	64.9	58.26	1
Self-employed	7.74	8.03	
Not employed	19.85	22.33	1
N	3839	28871	

Notes : Means are weighted. [1]NHANES mean significantly different than ATUS mean (p<.05).

Table 2. OLS Model Predicting Time Exercising with Demographic Characteristics and Interactions

	Base Models								Interacted Models							
	TSM				TCM				TSM				TCM			
	Coeff.		SE		Coeff.		SE		Coeff.		SE		Coeff.		SE	
Main Effects																
Gender [a]																
Women	-9.296	***	1.576		-3.378	+	1.898		-10.396	***	2.203		-1.124		1.951	
Men *(ref.)*																
Age																
19-24 (ref.)																
25-34	-5.334	+	2.653		-5.542	*	2.311		-5.703		4.131		-5.519		3.696	
35-44	-6.357	+	3.267		-6.207	*	2.529		-6.542		5.011		-5.991		3.962	
45-54	-8.101	**	2.789		-5.679	*	2.291		-9.67	*	3.977		-5.677		3.56	
55-64	-11.48	**	3.927		-6.832	*	2.617		-15.526	**	4.535		-7.774	+	3.944	
Race [ab]																
Non-Hispanic White *(ref.)*																
Non-Hispanic Black	-1.219		1.03		-1.084		0.949		-0.666		1.474		-0.293		1.317	
Hispanic	3.963		2.653		2.329		2.525		5.164		4.292		2.192		4.174	
Other	-6.779	*	2.678		-4.664	**	1.468		-9.028	*	3.512		-5.816	**	2.03	
Immigrant																
Born outside of the U.S.	-0.071		2.002		0.64		1.799		-1.542		2.726		-0.098		2.69	
Born in the U.S *(ref.)*																
Marital Status [ab]																
Married *(ref.)*																
Cohabiting	-3.299		3.386		-2.008		2.452		-7.113	*	3.308		-4.684	+	2.471	
Widowed/Divorced/Separated	-1.691		2.105		-0.972		1.831		-4.184	+	2.081		-3.379	*	1.65	
Never Married	2.449		2.465		1.633		2.578		-0.159		2.73		-1.773		2.405	
Educational Attainment [ab]																
Less than High School Degree	1.978		2.04		1.125		2.265		2.484		2.935		1.271		3.432	
High School Degree or Equivalent *(ref.)*																
Some College or Associate's Degree	-0.843		1.308		-0.793		1.232		-1.743		1.952		-1.672		1.763	
College Degree or Higher	3.18	*	1.402		4.055	**	1.147		3.055		2.288		4.681	*	1.767	
Family Income [b]																
Up to $34,999 *(ref.)*																
$35,000 to $74,999	-0.793		1.516		-0.526		1.346		-2.116		1.883		-1.781		1.598	
$75,000 and up	1.716		2.029		2.156		1.972		-0.792		2.164		-0.39		1.993	
Employment status																
Part-time	-0.232		1.897		1.251		1.35		-2.615		1.954		0.032		1.857	
Full-time *(ref.)*																
Self-employed	-0.27		2.589		2.219		1.631		-3.751	+	1.97		0.52		1.704	
Not employed	-3.995	+	2.142		-0.799		1.202		-6.097	*	2.422		-1.523		1.733	
Occupational Activity Classification																
Inactive *(ref.)*																
Active	6.233		3.695		0.81		2.214		10.803	***	2.838		3.182		2.334	
Weekday vs. Weekend																
Weekday *(ref.)*																
Weekend	-1.602		1.721		0.635		1.176		-3.875	*	1.639		-0.301		1.681	
BMI [ab]																
Under weight (<=18.5)	-2.41		3.315		-1.604		2.363		-3.852		5		-2.3		3.461	
Normal weight (18.5-24.9) *(ref.)*																
Overweight (25-29.9)	-3.918	**	1.196		-2.553	*	0.937		-4.18	*	1.799		-2.436		1.454	
Obese (30 or greater)	-10.12	***	2.222		-6.009	***	1.173		-13.04	***	1.549		-6.876	***	1.508	
Self-Reported Health																
Very Good/Excellent	1.172		1.506		0.417		1.56		-0.227		1.774		-1.419		1.4	
Good *(ref.)*																
Fair/Poor	-1.93		1.863		-0.867		1.875		-1.707		2.928		0.273		2.714	
ATUS Indicator [ab]																
ATUS	-13.913	***	0.78		5.149	***	0.724		-35.769	***	7.262		-4.916		7.233	
NHANES *(ref.)*																

Table 2 cont. OLS Model Predicting Time Exercising with Demographic Characteristics and Interactions

Interactions with ATUS								
Gender [d]								
Women					2.994	2.203	-6.316 **	1.951
Men *(ref.)*								
Age [c]								
19-24 (ref.)								
25-34					1.316	4.131	1.025	3.696
35-44					1.352	5.011	0.75	3.962
45-54					5.416	3.977	1.461	3.56
55-64					11.448 *	4.535	3.645	3.944
Race [c]								
Non-Hispanic White *(ref.)*								
Non-Hispanic Black					-1.544	1.474	-2.017	1.317
Hispanic					-2.341	4.292	0.611	4.174
Other					6.226 +	3.512	2.893	2.03
Immigrant								
Born outside of the U.S.					3.14	2.726	1.537	2.69
Born in the U.S *(ref.)*								
Marital Status [cd]								
Married *(ref.)*								
Cohabiting					10.36 **	3.308	7.751 **	2.471
Widowed/Divorced/Separated					7.522 **	2.081	6.497 ***	1.65
Never Married					7.575 **	2.73	8.797 ***	2.405
Educational Attainment								
Less than High School Degree					-1.907	2.935	-0.736	3.432
High School Degree or Equivalent *(ref.)*								
Some College or Associate's Degree					2.347	1.952	2.209	1.763
College Degree or Higher					0.461	2.288	-1.397	1.767
Family Income [d]								
Up to $34,999 *(ref.)*								
$35,000 to $74,999					3.713 +	1.883	3.461 *	1.598
$75,000 and up					6.61 **	2.164	6.307 **	1.993
Employment status [cd]								
Part-time					6.634 **	1.954	3.712 +	1.857
Full-time *(ref.)*								
Not employed					9.766 ***	1.97	5.323 **	1.704
Self-employed					7.255 **	2.422	2.638	1.733
Occupational Activity Classification [cd]								
Inactive *(ref.)*								
Active					-14.694 ***	2.838	-6.98 **	2.334
Weekday vs. Weekend [c]								
Weekday *(ref.)*								
Weekend					5.759 **	1.639	2.37	1.681
BMI [c]								
Under weight (<=18.5)					3.598	5	2.356	3.461
Normal weight (18.5-24.9) *(ref.)*								
Overweight (25-29.9)					1.508	1.799	-0.305	1.454
Obese (30 or greater)					8.896 ***	1.549	2.79 +	1.508
Self-Reported Health [d]								
Very Good/Excellent					4.359 *	1.774	5.579 ***	1.4
Good *(ref.)*								
Fair/Poor					-1.003	2.928	-2.879	2.714
Constant	43.187 ***	6.765	16.28 **	5.306	50.405 ***	7.262	19.236 *	7.233

Note: *p<0.05 **p<0.01 ***p<0.001, [a] Wald test p<.05 in TSM base model, [b] Wald test p<.05 in TCM base model, [c] Wald test p<.05 in TSM inteacted model, [d] Wald test p<.05 in TCM interacted model

Figure 1. Total Conditional Time Spent Exercising by Demographic Characteristics