Sleep duration: how well do self-reports reflect objective measures? The CARDIA Sleep Study

Diane S. Lauderdale[*], **Kristen L. Knutson**[*], **Lijing L. Yan**[†,‡], **Kiang Liu**[†], and **Paul J. Rathouz**[*]

[*]Department of Health Studies, University of Chicago, Chicago, Illinois

[†]Department of Preventive Medicine, Northwestern University, Chicago, Illinois

[‡]Department of Health Economics and Management, Guanghua School of Management, Peking University, Beijing, China

Abstract

Recent epidemiologic studies have found that sleep duration is associated with obesity, diabetes, hypertension and mortality. These studies have used self-reported habitual sleep duration, which has not been well validated. We model the extent to which self-reported habitual sleep reflects average objectively measured sleep. Eligible participants at the Chicago site of Coronary Artery Risk Development in Young Adults Study were invited to participate in a 2003-2004 ancillary sleep study; 82% (n=669) agreed. Sleep measurements collected in two waves included: 3-days of wrist actigraphy, a sleep log, and standard questions about usual sleep duration. Average measured sleep was 6 hours, and subjective reports averaged 0.80 hours longer than measured sleep. Subjective reports were not well calibrated, increasing on average by 31 minutes for each additional hour of measured sleep. Our model suggests that persons sleeping 5 and 7 hours over-reported, on average, by 1.3 and 0.3 hours respectively. Overall, there was a correlation of 0.45 between reported and measured sleep duration. The extent of overestimation, calibration and correlation varied by personal and sleep characteristics. Although asking about sleep duration seems uncomplicated, the correlation between self-reported and objectively-measured sleep in this population was moderate and systematically biased.

Recent epidemiologic studies have found that sleep duration is associated with obesity, diabetes, hypertension and mortality (1-13). These studies have in part been motivated by exciting findings from sleep laboratory studies that have demonstrated reduced sleep hours produce short-term metabolic and hormonal derangements, notably impaired glucose tolerance and increased appetite (14-16). Thus sleep duration has become a potentially important and novel risk factor for chronic disease. Sleep, though, is measured differently in experimental sleep laboratory studies than it is in most epidemiologic studies. In a sleep laboratory, hours available for sleep are carefully controlled, and sleep is precisely monitored through polysomnography. Many of the epidemiologic investigations to date have mined established cohorts that included a survey question such as "How many hours of sleep do you usually get a night (or when you usually sleep)?" (1). There has been little validation of such questions relative to more objective measures of sleep.

In this study we estimate the extent to which self-reports of usual sleep hours reflect an average of objectively measured sleep durations. We also examine the extent to which a single-day

Corresponding Author: Diane S. Lauderdale, Department of Health Studies, University of Chicago, 5841 S. Maryland Avenue MC 2007, Chicago, IL 60637, (lauderdale@uchicago.edu).

report of how much a person slept the previous night reflects measured sleep for that night. The study population includes healthy adults in early middle age.

Materials and Methods

Study sample

This is an ancillary study to an on-going prospective multi-center cohort, the Coronary Artery Risk Development in Young Adults (CARDIA) study. At recruitment in 1985-86, the CARDIA cohort was aged 18-30 and balanced by sex, race (black and white), and education. A more detailed study description has been presented elsewhere (17). This ancillary study includes participants from Chicago, one of four CARDIA sites. Non-pregnant participants in the CARDIA Year 15 clinical examination (2000-2001) were invited to participate in the sleep study in 2003 and 2004; 669 of 814 (82%) agreed to do so. Sleep study participants and non-participants had similar responses to questions about sleep asked in the 2000-2001 CARDIA interview: average self-reported sleep hours in the previous month had a mean of 6.5 hours for both groups, and the percentage reporting trouble falling asleep was 19 for participants and 20 for non-participants (18). Therefore, participants in the ancillary sleep study appear not to have self-selected by perceptions of sleep problems. All participants gave informed written consent; the protocol was approved by the institutional review boards of Northwestern University and the University of Chicago, and by the CARDIA steering committee.

Measurements

Sleep variables—Sleep data were collected in two waves about one year apart for each participant, between 2003 and 2005. In both waves sleep was measured using a wrist actigraphy monitor. Wrist actigraphy is an unobtrusive, objective method for identifying sleep periods. An actigraphy monitor (model AW-16, Mini Mitter, Inc, Bend, OR) looks like a wristwatch with a blank face. Using highly sensitive accelerometers, actigraphs digitally record an integrated measure of gross motor activity, which is analyzed to identify sleep periods. Wrist actigraphy has been compared to polysomnography – the "gold standard" for measuring sleep, demonstrating a correlation between subjects in sleep duration over 0.9 in healthy subjects (19). The mean absolute discrepancy ranges from 12-25 minutes among healthy non-elderly adults.

Consenting subjects were mailed actigraphy monitors and three standard sleep questionnaires (the Pittsburgh Sleep Quality Index, the Epworth Sleepiness Scale and the Berlin Questionnaire) (20-22) and were asked to wear the monitors for three nights, preferably Wednesday, Thursday and Friday. Participants then returned the monitor and questionnaires in a prepaid mailer. Actigraphy data were uploaded and sleep duration analyzed using manufacturer-supplied software. Sleep duration excludes periods of wakefulness during the night.

The actigraphy monitor has an event marker that may be pressed to mark specific times; participants were asked to push it each night when they began trying to fall asleep and again when they got out of bed each morning. The event marker does not start or stop data recording. Participants were also sent a sleep log to record bedtime and wake time each day ("Please report the time you get into bed and try to go to sleep ("Bedtime") and the time you got out of bed ("Wake Time") in the spaces provided. Write down the exact time, such as "11:37" am/ pm."). The sleep log provides backup data for bedtime and wake time when participants forgot to press the event marker. The bedtime and wake time are necessary to determine sleep duration, because the software only "looks for" sleep during that period. Otherwise, low motion periods during the day, or times when the watch is actually removed, could be counted as sleep.

Objective habitual sleep duration: To determine mean sleep duration for each subject at each wave, we calculated a weighted average of the weekday and weekend recordings. Some people wore the monitor a different three days or less than three days. We excluded people with only weekday or only weekend recordings. In Wave 1, 19 persons were dropped for inadequate data. For those with only one weekday recording (n=42), this value was repeated as a second weekday value, and for those (generally the same subjects) with two weekend values (n=35) we used the mean of these values. Each subject therefore had two weekday and one weekend actigraphy sleep duration measure. Mean sleep duration was weighted by day of week: 5/7 * ½ (weekday1 + weekday2) + 2/7 * (weekend) to obtain an objective measure of habitual sleep duration.

For the analysis of habitual subjective and objective sleep we only use data from Wave 1 because participants were mailed a report including their night-by-night actigraph-measured sleep duration after wave 1 participation. The report may have caused a learning effect and influenced subjective responses in wave 2.

Subjective habitual sleep duration: The Pittsburgh Sleep Quality Index includes these questions: "During the past month, how many hours of *actual sleep* did you get at night? (This may be different than the number of hours you spend in bed.) On week days? On week-ends?" (20). Week days and weekends were weighted (5/7 and 2/7) to yield subjective habitual sleep duration for each wave.

Subjective nightly sleep duration (for each night of actigraphy recording): In Wave 2, participants were also asked on the sleep log, "What is your best estimate of how much actual sleep you got each night?"

The following sleep, health and sociodemographic variables were all dichotomized for stratified analyses, so that we could examine how and test whether the stratifying variables affect the association between objective and subjective sleep. We do not incorporate all of these potential effect modifiers in a single multivariate model because of the complexity of including (and interpreting) so many terms all interacted with objectively measured sleep while simultaneously accounting for measurement error.

Sleep efficiency: Sleep efficiency is a ratio of the time sleeping divided by time in bed (after one begins trying to fall asleep). The event marker (or backup log) was used to calculate the time in bed. Sleep efficiency was dichotomized at 80%.

Sleep variability: We calculated the difference between the nights with the longest and shortest sleep durations during the three nights. Persons with more than a two-hour difference were considered to have high sleep variability.

Sleepiness: The Epworth Sleepiness Scale includes 8 items and assesses the general level of daytime sleepiness. Scores range from 0-24 where higher scores indicate greater sleepiness. Following the developer's suggestion, a score greater than 10 was classified as high daytime sleepiness. (21)

Apnea risk: The Berlin Questionnaire was used to identify respondents at high risk of sleep apnea. A participant is classified as high risk if he/she has two of the three following conditions: (1) loud or frequent snoring or frequent breathing pauses (2) frequently tired after sleeping or during waketime or having fallen asleep while driving, (3) high blood pressure or BMI > 30 kg/m^2. (22)

Sociodemographic and health variables—These variables were collected during the CARDIA Year 15 interview in 2000-2001.

Race and *sex* were collected at cohort initiation and verified in 2000-2001. All participants are black or white.

Age at the time of actigraphy recording is dichotomized at the sample mean, 42 years.

College Graduates were identified using a question about highest education obtained.

Household income in 2000-2001 was dichotomized as low (<$35,000/year) versus high.

Obesity. Body Mass Index (BMI) was calculated by dividing measured weight (kg) by height squared (m^2). Obesity was defined as a BMI of 30 or greater.

Depression was measured using the Center for Epidemiological Studies–Depression scale. (23) Following prior use, persons with a score of 16 or higher were categorized as having a high depression score.

Self-rated health was a five-level response: poor, fair, good, very good and excellent. This was dichotomized as fair or poor versus good, very good or excellent.

Analysis—To compare subjective and objective measures of sleep, we focus on three aspects of the subjective-objective relationship: *bias*, *calibration* and *discrimination* of subjective sleep treating objective sleep as the gold standard. These constructs are naturally operationalized in a linear regression model of subjectively measured sleep on objectively measured sleep. *Bias* captures the degree to which, on average, subjects over or under-estimate sleep; it is measured via the regression intercept. Were there no bias, the intercept would be 0 hours. We report the intercept at the average of 6 hours measured sleep. *Calibration*, the sensitivity of the subjective response to variation in the objective response, is captured by the regression slope. Perfectly calibrated measures would have a slope of one: one additional hour of measured sleep would predict, on average, one additional hour of subjective report. Finally, *discrimination* measures the degree to which individuals with higher objective measures also tend to be those with higher subjective measures, regardless of bias or calibration. It is captured via the model r-squared value (or its square root, the correlation).

We carried out two analyses of the subjective-objective relationship. The first focused on habitual sleep (past 30 day average) while the second examined sleep for a single night.

In an ideal design, the first analysis would regress self-reported habitual sleep on a 30-day average of objectively-measured sleep. Instead, we only have a 3-day weighted average of measured sleep. We expect some error in using this 3-day average in place of the 30-day average, yielding an errors-in-variables problem. Ignoring the error in right hand side variables is known to produce attenuation bias in the regression coefficients (24). The error variance in the regressors is often quantified via the *reliability* coefficient, i.e., the ratio of the variance of the true 30-day average sleep duration to the variance of the 3-day weighted average sleep duration. A reliability of one indicates no measurement error, and the reliability tends toward zero as the measurement error becomes more important. Fortunately, if one can estimate the error variance or, equivalently, the reliability coefficient, then it is straightforward to correct the attenuation bias (25). We treat each subject's set of 3-days of objectively-measured sleep as a stratified random sample from 30 days on that same subject, with a sample size of two in the weekday stratum and of one in weekend stratum. The error variance of the 3-day weighted average of objective sleep relative to the 30-day average is estimated as a weighted function of stratum-specific variances, just as in stratified survey samples (26, Section 3.2). The

weekday variance is estimated as the average across all subjects of the realized within-subject sample variances for the two weekdays. Since we only have one observation per subject on weekends, weekend variance was assumed to be equal to weekday variance. Using the estimated error variance of the 3-day weighted average, and the total between-subject variance of the 3-day weighted average, we were able to estimate the reliability coefficient for the 3-day weighted average as a surrogate for the 30-day average. This reliability coefficient was then fed into errors-in-variables regression models.

We then performed errors-in-variables regression of subjective habitual sleep duration on 3-day weighted average objective sleep duration. Standard errors were obtained via the bootstrap (27), which accounted for uncertainty in estimation of the reliability coefficient. Confidence intervals and hypothesis tests for the r-squared values were conducted on the Fisher's *z*-transformation of *r*.

We conducted analyses on the full sample, and then stratified by variables hypothesized to modify the association between subjective and objective sleep duration. These are sociodemographic (sex, race, education, income, age), health (obesity, depression, self-rated health) and sleep (apnea risk, sleepiness, sleep efficiency, sleep variability) variables.

The second analysis regresses self-reported sleep for a single night on measured sleep duration for that same night. To examine whether people accurately perceive sleep in a single night, the sleep log was modified in the second wave of sleep recordings, and participants were asked to record each morning how much they actually slept the previous night. This sleep estimate is likely to be more accurate than one made outside our study setting for two reasons: participants had received the sleep report from wave 1, and participants were concurrently keeping a sleep log. Since subjects had three sequential nights of measured sleep, we use generalized estimating equations to fit the models and robust (empirical) variance estimators for confidence intervals and hypothesis tests (28, Ch.8). We use the independence correlation structure to avoid bias arising from objective sleep on a given night predicting subjective sleep on another night within the same subject. (29)

We carried out a sensitivity analysis to check whether inaccurately recorded bedtimes and wake times could have influenced our main findings, since the software only looks for sleep between these two times. We examined the actigraphy data right before and after each main sleep period to see whether there was inactivity that the software would have interpreted as sleep were it in the interval between bedtime and wake time -- even though such inactivity is not necessarily sleep. These records were removed and analyses rerun.

Analyses were carried out in Stata version 9. (30) Errors-in-variables models were fitted with the "eivreg" function; generalized estimating equations models with the "regress" function while clustering on each subject. Bootstrap standard errors were computed with the Stata bootstrap utility.

Results

We excluded 22 of 669 participants from the main analysis for these reasons: 19 lacked either weekday or weekend recordings, 1 did not complete the sleep questionnaires, 1 appeared to have removed the actigraph during the night and one outlier for whom almost no sleep was recorded. Thus the final sample for wave 1 analysis comprised 647 subjects (table 1). Mean measured sleep duration was 6.06 hours. Mean self-reported habitual sleep was 6.83 hours, and only 17 percent reported less than the measured mean.

For habitual sleep in wave 1, the bias at the mean of 6 hours measured sleep was 0.80 hours (48 minutes), with subjective reports longer than measured sleep (table 2). The calibration,

represented by the beta coefficient, was substantially less than one: for each additional hour of mean sleep recorded, report of habitual sleep increased, on average, by 31 minutes. Mean measured sleep explained 20 percent of the variation ($r^2 = 0.20$) in reported habitual sleep, a correlation of 0.45. Combining the effects of bias and calibration, persons who slept 5 hours reported, on average, 6.29 hours of sleep, and persons who slept 7 hours reported 7.31 hours.

Bias varied little by sex, education, income or sleep variability, but did vary significantly by several demographic, health and sleep variables (table 2). The bias was closer to 0 for blacks, the obese, those with high depression scores, high apnea risk, high sleepiness, and high sleep efficiency. The stratification that made the greatest difference in bias was apnea risk, with low risk persons overestimating sleep by an average of 54 minutes, and high risk persons overestimating sleep by only 10 minutes.

Calibration did not vary significantly for most of the stratification variables, but was better (closer to one) for those with higher sleep efficiency (Table 2). An additional hour of measured sleep corresponded to 47 minutes more of reported sleep for those with higher sleep efficiency, but 25 minutes more of reported sleep for those with low sleep efficiency. Calibration was also better (although the comparisons were not statistically significant) for those with higher incomes, more education and older age. Similarly, the r^2 was highest (.29 to .34, corresponding to correlations between .54 and .58) for whites, persons older than the mean, those with more education, more income, and higher sleep efficiency. The correlation was lower (between .20 and .40) for blacks, persons younger than the mean, without a college degree, with low income or low sleep efficiency. Sleep variability did not significantly affect bias, calibration or discrimination. Both calibration and correlation were close to 0 for persons in fair/poor health. Since less than 10% of the sample was in this category, only the correlation contrast attained statistical significance.

Because of lower participation in wave 2, the final sample of single-night sleep was 615 subjects. For a single night, the bias was 0.63 hours (38 minutes), with subjective reports longer than measured sleep (data not shown). For each additional hour of sleep recorded, the report of sleep duration increased by 35 minutes. Measured sleep explained 36 percent of the variation in reported sleep for a single night, a correlation of 0.60.

The sensitivity analysis identified 216 actigraphic records where sleep for at least one night would have been 30 minutes longer if the period before the recorded bedtime and after the recorded wake time were scanned for inactivity that resembled sleep. However, repeating the main analysis of habitual objective-subjective sleep without these records did not improve the estimates of bias, calibration or discrimination. The bias of the remaining records was 0.86 hours, the calibration was .45 and the discrimination was .17 (data not shown). Stratified results were also very similar to the full sample, although some comparisons were no longer statistically significant, consistent with smaller sample sizes.

Discussion

We found a correlation between self-reported and objectively measured sleep duration of 0.45, which is generally be considered a "moderate" correlation. The correlation is significantly lower for some subpopulations. We found evidence of systematic errors in both the mean and the calibration. The subjective mean is almost an hour greater than the measured mean at the average of 6 hours of measured sleep, and each additional hour of objective sleep is reflected by only a half hour of additional subjective reported sleep. Overall, 20 percent of the variation in subjective report of habitual sleep is explained by variation in measured sleep. We thought of two explanations for why it might be difficult to estimate usual sleep duration. One possibility is that people cannot accurately report how much they sleep on a single night; the

other possibility is that high night-to-night variability makes it difficult to integrate information over 30 nights. Our data support the former. We did not find that those with smaller night-to-night differences in sleep were less biased or better calibrated than persons with greater variability. We did find that single-night estimates were only a little more accurate than the reports of habitual sleep – even though subjects concurrently kept a sleep log.

Logically, another possible factor contributing to the bias (although not the low calibration) could be a problem with the actigraphy, especially systematic underestimation of sleep duration. While our study did not include an internal validation of actigraphy, many prior studies have compared actigraphy with concurrent polysomnography, the gold standard. In a 2003 comprehensive review, correlations between actigraphy and polysomnography for duration were over 0.9 in healthy adults. (31) Correlations in clinical studies of persons with sleep disorders are lower (most between 0.7 and 0.9), and actigraphy seems to systematically *overestimate* sleep duration for insomniacs because still awakenings are counted as sleep. (32,33) Most studies of healthy individuals have not found systematic bias, but when they have it is towards overestimated sleep duration. (34) Actigraphy has recently been added to several large population-based cohorts, including the Study of Women's Health Across the Nation (SWAN), (35) which focuses on the menopausal transition, and the Study of Osteoporotic Fractures, (36) which includes older women.

There have been very few prior validation studies of self-reported habitual sleep. One population-based study validated a usual sleep hour question by comparing it to an average of daily self-reported sleep durations, kept in a one-week log (11). The Spearman correlation was 0.79. Both were collected in the same mailing. Our finding that self-reports of single nights have similar bias and calibration as reports of habitual sleep suggests limitations to this validation.

Another study compared several self-reported measures of sleep quality and quantity to actigraphy, among postmenopausal women experiencing hot flashes. (37) Mean actigraph sleep over seven nights averaged 6.3 hours, and mean self-reported sleep averaged 6.6 hours. Women who under-estimated sleep duration were more likely to report low quality sleep. They found strong associations between poor self-reported sleep quality (often unconfirmed by actigraphy) and measures of psychological and somatic distress, consistent with complex factors influencing subjective reports about both sleep quantity and quality.

Our study has several data limitations. First, this is one cohort and one age group. Since we found many factors affected bias and calibration, it seems likely that other populations might differ in bias, calibration and correlation. Second, there is no perfect way to measure sleep duration without disrupting routine. Actigraphy does not perturb normal sleep habit as there appears to be no "first-night effect." (31) However, prior evidence that the accuracy of actigraphy is worse in insomniacs could be a factor in some of our stratified contrasts. Third, we only have 3 nights of actigraphy, which we treat as a random sample from 30. However, we have used measurement error methods to correct for bias in measurement. Fourth, we only have data about single-night self-report from wave 2 after participants received a summary of their sleep from wave 1 and while they were keeping a log, both of which seem likely to increase the correlation. Finally, several of the stratification variables were measured two to three years before sleep recordings, and may have changed; and we do not have an apnea diagnosis, just risk approximated using the Berlin Questionnaire.

One explanation for these findings could be that people are generally not sure how much they sleep, and when pressed to give a response on a survey tend to answer what they believe to be how much adults in general sleep – our modal answers were 7 hours for weeknights and 8 hours for weekends. Since most actually sleep less than that, it is generally an overestimate,

but it is less of an overestimate for people who sleep more and more of an overestimate for people who sleep less. However, persons with health problems (such as depression or obesity) or who feel tired, might suspect they sleep less than the norm, regardless of their actual measured sleep. Thus their overestimates are smaller. In fact, people who report fair or poor health (relatively rare among this community-based sample of persons in their forties) actually have no significant correlation between measured and reported sleep, and report on average shorter sleep hours than those with better self-rated health. The implications are important. If other populations of study participants behave similarly, then there may be significant associations between self-reported sleep duration and health that are not due to actual sleep duration, but there may also be true associations that are masked.

Many facets of sleep could be associated with health: apnea, sleep stages, duration and disruption, and subjective perceptions of quality, quantity and sleepiness; all pose measurement challenges for epidemiologists. While actigraphy provides an adequately accurate measure of duration and fragmentation, it leaves important sleep characteristics unmeasured. Industry is active in new instrument development, and there may be better options soon. For survey data collection, cognitive testing in diverse populations about how respondents arrive at an answer to a question about usual sleep duration might suggest additional approaches to asking these questions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Literature cited

1. Gangwisch JE, Heymsfield SB, Boden-Albala B, Buijs RM, Kreier F, Pickering TG, Rundle AG, Zammit GK, Malaspina D. Short sleep duration as a risk factor for hypertension: analyses of the first National Health and Nutrition Examination Survey. Hypertension 2006 May;47(5):833–9. [PubMed: 16585410]

2. Yaggi HK, Araujo AB, McKinlay JB. Sleep duration as a risk factor for the development of type 2 diabetes. Diabetes Care 2006 Mar;29(3):657–61. [PubMed: 16505522]

3. Knutson KL. Sex differences in the association between sleep and body mass index in adolescents. J Pediatr 2005 Dec;147(6):830–4. [PubMed: 16356441]

4. Gangwisch JE, Malaspina D, Boden-Albala B, Heymsfield SB. Inadequate sleep as a risk factor for obesity: analyses of the NHANES I. Sleep 2005 Oct 1;28(10):1289–96. [PubMed: 16295214]

5. Mallon L, Broman JE, Hetta J. High incidence of diabetes in men with sleep complaints or short sleep duration: a 12-year follow-up study of a middle-aged population. Diabetes Care 2005 Nov;28(11): 2762–7. [PubMed: 16249553]

6. Gottlieb DJ, Punjabi NM, Newman AB, Resnick HE, Redline S, Baldwin CM, Nieto FJ. Association of sleep time with diabetes mellitus and impaired glucose tolerance. Arch Intern Med 2005 Apr 25;165 (8):863–7. [PubMed: 15851636]

7. Taheri S, Lin L, Austin D, Young T, Mignot E. Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. PLoS Med 2004 Dec;1(3):e62. [PubMed: 15602591] Epub 2004 Dec 7

8. Amagai Y, Ishikawa S, Gotoh T, Doi Y, Kayaba K, Nakamura Y, Kajii E. Sleep duration and mortality in Japan: the Jichi Medical School Cohort Study. J Epidemiol 2004 Jul;14(4):124–8. [PubMed: 15369129]

9. Hasler G, Buysse DJ, Klaghofer R, Gamma A, Ajdacic V, Eich D, Rossler W, Angst J. The association between short sleep duration and obesity in young adults: a 13-year prospective study. Sleep 2004 Jun 15;27(4):661–6. [PubMed: 15283000]

10. Kripke DF, Garfinkel L, Wingard DL, Klauber MR, Marler MR. Mortality associated with sleep duration and insomnia. Arch Gen Psychiatry 2002 Feb;59(2):131–6. [PubMed: 11825133]

11. Patel SR, Ayas NT, Malhotra MR, White DP, Schernhammer ES, Speizer FE, Stampfer MJ, Hu FB. A prospective study of sleep duration and mortality risk in women. Sleep 2004 May 1;27(3):440–4. [PubMed: 15164896]

12. Ayas NT, White DP, Manson JE, Stampfer MJ, Speizer FE, Malhotra A, Hu FB. A prospective study of sleep duration and coronary heart disease in women. Arch Intern Med 2003 Jan 27;163(2):205–9. [PubMed: 12546611]

13. Ayas NT, White DP, Al-Delaimy WK, Manson JE, Stampfer MJ, Speizer FE, Patel S, Hu FB. A prospective study of self-reported sleep duration and incident diabetes in women. Diabetes Care 2003 Feb;26(2):380–4. [PubMed: 12547866]

14. Spiegel K, Leproult R, Van Cauter E. Impact of sleep debt on metabolic and endocrine function. Lancet 1999;354:1435–1439. [PubMed: 10543671]

15. Spiegel K, Leproult R, L'Hermite-Baleriaux M, et al. Leptin Levels Are Dependent on Sleep Duration: Relationships with Sympathovagal Balance, Carbohydrate Regulation, Cortisol, and Thyrotropin. J Clin Endocrinol Metab 2004;89:5762–5771. [PubMed: 15531540]

16. Spiegel K, Tasali E, Penev P, et al. Sleep Curtailment in Healthy Young Men Is Associated with Decreased Leptin Levels, Elevated Ghrelin Levels and Increased Hunger and Appetite. Ann Intern Med 2004;141:846–50. [PubMed: 15583226]

17. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol 1988;41:1105–16. [PubMed: 3204420]

18. Lauderdale D, Knutson K, Yan L, et al. Objectively measured sleep characteristics among early middle-aged adults: The CARDIA Study. Am J Epidemiol 2006;64:5–16. [PubMed: 16740591]

19. Jean-Louis G, von Gizycki H, Zizi F, et al. The actigraph data analysis software: I. A novel approach to scoring and interpreting sleep-wake activity. Percept Mot Skills 1997;85:207–16. [PubMed: 9293579]

20. Buysse DJ, Reynolds CF III, Monk TH, et al. The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research. Psychiatry Res 1989;28:193–213. [PubMed: 2748771]

21. Johns MW. A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. Sleep 1991;14:540–545. [PubMed: 1798888]

22. Netzer N, Stoohs R, Netzer C, et al. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 1999;131:485–491. [PubMed: 10507956]

23. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Measurement 1977;1:385–401.

24. Carroll, RJ.; Ruppert, D.; Stefanski, LA. Measurement Error in Nonlinear Models. Chapman and Hall; London: 1995.

25. Kmenta, J. Elements of Econometrics. Vol. 2nd. University of Michigan Press; AnnArbor, MI: 1997.

26. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. Chapman and Hall; New York: 1993.

27. Kish, L. Survey Sampling. John Wiley and Sons; New York: 1965.

28. Diggle, PJ.; Heagerty, P.; Liang, KY.; Zeger, SL. Analysis of Longitudinal Data. Vol. 2nd. Oxford University Press; Oxford; 2002.

29. Pepe MS, Anderson GA. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Communications in Statistics – Simulation 1994;23:939–51.

30. StataCorp.. Stata Statistical Software: Release 9. College Station TX: StataCorp LP.; 2005.

31. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003;26:342–92. [PubMed: 12749557]

32. Lichstein KL, Stone KC, Donaldson J, et al. Actigraphy validation with insomnia. Sleep 2006;29:232–239. [PubMed: 16494091]

33. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med 2001;2:389–96. [PubMed: 14592388]

34. de Souza L, Benedito-Silva AA, Pires ML, Poyares D, Tufik S, Calil HM. Further validation of actigraphy for sleep studies. Sleep 2003;26:81–5. [PubMed: 12627737]

35. Hall M, Kravitz HM, Gold E, Sowers MF, Matthews KA, Buysse DJ, Sanders MA, Bromberger JT, Owens JF. Sleep during the menopausal transition in a multi-ethnic cohort: Feasibility and preliminary results. Sleep 2005;28(Abstract Supplement):A119.

36. Blackwell T, Yaffe K, Ancoli-Israel S, Schneider JL, Cauley JA, Hillier TA, Fink HA, Stone KL, Study of Osteoporotic Fractures Group. Poor sleep is associated with impaired cognitive function in older women: the study of osteoporotic fractures. J Gerontol A Biol Sci Med Sci 2006;61:405–10. [PubMed: 16611709]

37. Regestein QR, Friebely J, Shifren JL, Scharf MB, Wiita B, Carver J, Schiff I. Self-reported sleep in postmenopausal women. Menopause 2004;11:198–207. [PubMed: 15021450]

Table 1

Study sample characteristics for wave 1 of the sleep study, Chicago site of CARDIA, 2003-2004.

N	647
Mean (standard deviation)	
Mean age in years	42.9 (3.6)
Mean objective sleep duration in hours	6.06 (1.16)
Mean subjective sleep duration in hours	6.83 (1.11)
Proportion	
Female	0.58
Black	0.44
College graduate	0.51
Income ≥ $35,000/year	0.67
Obese	0.33
High depression score	0.17
Fair or poor self-rated health	0.07
Low sleep efficiency	0.32
High risk of apnea	0.14
High sleepiness	0.23
High sleep variability	0.37

Table 2

Linear regression models predicting reported habitual sleep hours by average sleep duration measured using wrist actigraphy. Regression is a modified version of errors-in-variables regression. Data collected in 2000-2004 from the Chicago site of CARDIA.

Stratification variable	Reliability[†]	Bias (at 6 hours) Intercept (hours)	95% CI	Calibration Beta	95% CI	Discrimination r²	95% CI	Predicted reported sleep hours at 5 hours measured sleep	7 hours measured sleep
Full Sample (n=647)	0.70	0.80	0.71, 0.88	0.51	0.35, 0.67	0.20	0.11, 0.30	6.29	7.31
Male	0.75	0.82	0.71, 0.92	0.46	0.28, 0.63	0.21	0.10, 0.33	6.36	7.27
Female	0.62	0.75	0.60, 0.90	0.59	0.29, 0.89	0.19	0.06, 0.34	6.16	7.34
Black	0.61	0.63*	0.47, 0.79	0.38	0.11, 0.65	0.07**	0.01, 0.19	6.25	7.01
White	0.73	0.88*	0.77, 0.99	0.54	0.37, 0.70	0.29**	0.18, 0.42	6.34	7.42
Age ≤ 42 years	0.72	0.88	0.74, 1.02	0.38	0.19, 0.57	0.11*	0.03, 0.22	6.50	7.26
Age > 42 years	0.69	0.74	0.64, 0.85	0.62	0.41, 0.83	0.29*	0.16, 0.42	6.12	7.36
Non-obese	0.70	0.90**	0.80, 1.00	0.47	0.31, 0.63	0.19	0.10, 0.29	6.43	7.37
Obese	0.69	0.61**	0.45, 0.77	0.54	0.19, 0.89	0.19	0.05, 0.36	6.07	7.15
Low depression score	0.71	0.86**	0.76, 0.96	0.50	0.33, 0.67	0.20	0.11, 0.30	6.36	7.36
High depression score	0.65	0.52**	0.32, 0.72	0.47	0.11, 0.83	0.16	0.02, 0.36	6.05	6.99
Good-excellent	0.70	0.80	0.71, 0.88	0.57	0.42, 0.72	0.24*	0.15, 0.33	6.22	7.37
Poor/Fair	0.67	0.54	0.11, 0.97	0.06	-0.60, 0.73	0.00*	0.00, 0.17	6.48	6.60
No college degree	0.69	0.73	0.61, 0.85	0.43	0.22, 0.63	0.12*	0.04, 0.23	6.30	7.16
College degree	0.69	0.83	0.71, 0.95	0.59	0.37, 0.81	0.30**	0.16, 0.45	6.24	7.42
Low income	0.68	0.76	0.59, 0.94	0.36	0.11, 0.61	0.08**	0.01, 0.20	6.41	7.12
High income	0.69	0.75	0.64, 0.85	0.67	0.47, 0.87	0.34**	0.22, 0.40	6.08	7.42
Low apnea risk	0.68	.090**	0.80, 1.00	0.48	0.33, 0.63	0.18	0.11, 0.28	6.42	7.38
High apnea risk	0.75	0.17**	-0.10, 0.44	0.46	0.10, 0.82	0.17	0.02, 0.39	5.71	6.63
Low sleepiness	0.65	0.87**	0.76, 0.97	0.48	0.30, 0.66	0.16	0.08, 0.26	6.39	7.35
High sleepiness	0.77	0.56**	0.35, 0.76	0.47	0.19, 0.75	0.18	0.04, 0.35	6.08	7.03
Low sleep efficiency	0.67	0.93*	0.71, 1.15	0.42*	0.16, 0.67	0.11*	0.03, 0.23	6.51	7.34
High sleep efficiency	0.62	0.60*	0.45, 0.76	0.78*	0.54, 1.03	0.32*	0.18, 0.47	5.82	7.39
Low sleep variability	0.90	0.78	0.68, 0.87	0.52	0.39, 0.66	0.24	0.14, 0.35	6.25	7.29
High sleep variability	0.46	0.83	0.68, 0.99	0.50	0.00, 1.00	0.14	0.01, 0.36	6.33	7.36

[†] Reliability of 3 day weighted average as a measure of true 30-day average

* p<0.05 for comparison of intercept, beta coefficient or r^2.

** p<0.01 for comparison of intercept, beta coefficient or r^2.