Sleep and Aging in Population Research

Diane S. Lauderdale, PhD
Professor (Epidemiology) and Chair
Department of Public Health Sciences
University of Chicago

July 24, 2016

INTRODUCTION AND OVERVIEW

Sleep is very important to older adults, and research studies about sleep, particularly those that identify problems with sleep and consequent ill health, receive significant media attention. However, making sure that research findings about sleep are accurate and informative is quite challenging. Incorporating sleep into studies of health and aging is complicated both conceptually and practically. Sleep is not a single entity, and no single approach to measuring sleep can capture all the dimensions that may be relevant for health, nor is there a summary number that describes it. Sleep is multi-faceted with behavioral, social, psychological, and biological components. Adding to the complexity of learning about sleep from individual respondents is the problem that they do not observe many aspects of sleep. Thus, the reporting challenges for sleep are different and perhaps even greater than health behaviors well recognized to be challenging for individuals to report accurately such as diet and physical activity. In addition, sleep research has multiple disciplinary origins, and they approach the conceptualization and measurement of sleep in different ways. Psychological approaches often rely on instruments with multiple questions that are designed to tap into latent constructs; epidemiological approaches query quantifiable characteristics; and clinical sleep medicine focuses on diagnosing defined sleep disorders or "diseases."

This paper will first sketch out why sleep is so complicated by describing some of the dimensions of sleep that may be relevant for population studies and the ways sleep has been most often measured in population research. As with any variable employed in research about groups of respondents, random noise is much less of a problem for inferences than is biased measurement. Finally, there will be a summary of outstanding questions about sleep and health that studies of older adults might investigate.

DIMENSIONS OF SLEEP

Sleep is a *behavior*: individuals set aside certain hours for sleep. There is variation both within individuals and between individuals. The times set aside for sleep—that is, when it commences and how long it lasts—may be the same every 24-hour period, vary day by day, or, more systematically, vary between workdays/weekdays and non-workdays/weekends. There may be one sleep period in 24 hours, or more than one (i.e., naps). The circadian timing of sleep also varies between individuals, some staying up late at night ("owls") and some waking early in the morning ("larks"). Typically, circadian timing shifts with aging, to an earlier bedtime and waking time.

Sleep has a *setting*, both within the home and within a neighborhood. The neighborhood may be safe or dangerous, quiet or noisy; the bedroom/sleep space may be light or dark, hot or cold, damp or dry, odiferous or fresh, solitary or in company. A bed may be comfortable or not. The room may be the same every night or vary. The quality of the setting is likely to have strong economic correlates.

Then there is the *actual sleep* that occurs, which depends on the behavior and (probably) the setting. Sleep duration has been the most frequently examined dimension of sleep in health research, with strong longitudinal associations frequently reported between hours of sleep, and mortality and chronic disease incidence, such as diabetes and stroke. Often shorter sleep duration is interpreted as a "choice" in the sleep literature and in media coverage, but for many older adults, shorter sleep is not a behavioral choice, but rather an inability to initiate or maintain sleep. Below we will consider in detail how sleep duration is measured.

Sleep has many measurable features. Ones of which the individual is essentially unaware include sleep architecture, REM cycling, waking periods, and movement. Sleep architecture can be discerned from an electroencephalogram (EEG) and is the pattern of sleep cycles that alternate between REM (rapid eye movement) sleep and non-REM sleep. Non-REM sleep is further characterized as deep slow-wave sleep (stage 3, previously stages 3 and 4) versus lighter sleep (stages 1 and 2). Individuals typically cycle through these stages several times during a sleep period with waking more likely to occur during REM periods than other stages. In REM cycling, there is variation in the regularity of the cycling pattern between and within individuals, the duration and number of cycles, and how they change throughout sleep. In addition, there are waking periods that may occur, momentary or sustained, of which the individual may not be aware. There is also variation in how much individuals move while asleep.

Some characteristics of sleep are considered pathological: *sleep disorders*. A medical specialization (Sleep Medicine) arose in the wake of the dissemination of polysomnography in the 1970s to diagnose and treat them. Insomnia is the most prevalent sleep disorder, with 30 percent or more of older adults reporting some symptoms of insomnia.[1,2] A clinical insomnia diagnosis is based on patient-reported problems and not laboratory tests. The clinical definitions differ slightly between the *Diagnostic and Statistical Manual of Mental Disorders* and *International Classification of Diseases,* but the key features are similar: reported difficulty initiating or maintaining sleep (specifically falling asleep, staying asleep, or waking up too early) and not feeling rested in the daytime such that it interferes with activity. Four questions that tap into the definitional facets of insomnia are often included in surveys of older adults, and respondents report high prevalence of specific components of insomnia such as trouble falling asleep.

Other sleep disorders are less common in the general population and are diagnosed through medical history combined with laboratory tests. Obstructive sleep apnea (OSA) is discontinued breathing and resulting oxygen desaturation during sleep. The

symptoms are typically extreme sleepiness during the day and snoring or gasping for breath at night (which may only be known if there is a bed partner). It is to a large extent an anatomical problem and having fat around the neck (and thus obesity) are important risk factors for OSA, although slender individuals may also have OSA. Individuals are generally unaware of having OSA unless they have had polysomnography in a sleep clinic to diagnose it, but are likely to report daytime sleepiness and know about their snoring. Snoring has high sensitivity for OSA but low specificity, that is people with PSA generally snore but lots of snorers do not have OSA. Among less common sleep problems, the most prevalent is probably restless leg syndrome, which is involuntary leg motions that interfere with sleep. Though it is implicitly included in questions about overall sleep quality or "trouble sleeping," restless leg syndrome is not specifically included in general surveys, although prevalence may be as high as five percent of adults.[3]

A less clearly defined dimension of sleep is "quality," which may encompass sleep problems or disorders or other features of sleep, including how rested the individual feels in the morning. Sleep quality is an imprecise term, but in sleep research it often specifically refers to the score on the widely used Pittsburgh Sleep Quality Instrument (PSQI). The PSQI, developed by Daniel Buysse and colleagues, includes 19 self-rated questions and yields a global score with a cutpoint for poor quality sleep. The seven components are self-rated sleep quality ("very good" to "very bad"), sleep latency (how long it takes to fall asleep), sleep duration, habitual sleep efficiency (the proportion of time in bed spent sleeping), sleep disturbances (e.g., "bad dreams," "pain," "get up to use the bathroom"), use of sleeping medications, and daytime dysfunction (i.e., "staying awake while driving, eating meals, or engaging in social activity" and "keep up enough enthusiasm to get things done "), during the past month.[4] Probably due to its length, the full PSQI has not been used in omnibus surveys where sleep is not a primary focus of investigation. Other uses of the term "sleep quality" may just refer to a single self-rated question, insomnia symptoms, or features of sleep consolidation or efficiency measured through survey questions (such as those in the PSQI), or objectively measured using actigraphy or polysomnography (described below).

MEASUREMENT OF SLEEP IN POPULATION RESEARCH

Although technologic innovation has been expanding options and will continue to do so in the near future, at present, data collection about sleep is primarily accomplished in population research through survey questions and increasingly by accelerometers/actigraphy. Sleep logs are not generally used in population research but could be; they are often used in clinical sleep medicine. A few population studies with sizable numbers of participants (hundreds or a few thousand) with a primary or secondary focus on OSA have used various lab and home measurement approaches to diagnose OSA, or oxygen desaturation during sleep, but OSA measurement is not a focus of this paper.

POLYSOMNOGRAPHY (very briefly)

An often unacknowledged challenge in figuring out how to evaluate the accuracy and bias of the more feasible types of data collection, e.g., survey questions and actigraphy, is that for many aspects of sleep there really is no gold standard against which to measure them. Polysomnography (PSG) is often referred to as the gold standard for measuring sleep, but that is almost a tautology: when it was discovered that electrical activity in the brain could be used to differentiate different types of sleep, sleep was in a sense redefined to be the patterns observed through electroencephalograms. Sleep medicine as a certified medical specialty is tightly linked to the proliferation of sleep clinics with PSG that began in the 1970s, as well as pressure to control competition.[5] Full PSG requires over 20 electrodes measuring electrical brain activity, eye movements, muscle activity, heart rhythm, breathing, and oxygen saturation. However, while PSG is the definitive way to identify sleep stages and REM sleep, and to diagnose OSA and several sleep disorders, it is not very useful for measuring the dimensions of sleep that have been the focus in population research: sleep duration and insomnia, the most common sleep disorder. For both, the intrusiveness of the test often interferes with the subject's normal sleep behavior, such that sleep duration and difficulty falling or staying asleep (insomnia symptoms) may not represent the subject's typical experience when not attached to the sensors. The primary way that actigraphy is validated is to compare it to concurrent PSG during a single night. The (rarely acknowledged) problem with this is that it is unclear whether individuals sleep and move similarly when attached to all the PSG monitors to how they sleep at home. For example, subjects may be more likely to sleep on their backs with PSG monitors than they would otherwise or to stay still when they wake up. Thus, the many small clinic-based validation studies of sleep time for actigraphy, which generally find a reasonably good concordance with sleep time from concurrent PSG (correlation around .90[15]) may not represent how an actigraph works for someone who is not attached to PSG monitoring.

While PSG in a sleep lab is only carried out in research where OSA is the primary object of study, because of expense and respondent burden, there are somewhat less intrusive ways to measure OSA and even sleep stages, in the home. The Sleep Heart Health Study and the Multi-Ethnic Study of Atherosclerosis carried out an in-home version of PSG .[6] Other in-home measurements of sleep include the Apnea Risk Evaluation System (ARES) Unicorder for sleep apnea, which is a self-applied, portable monitoring instrument that evaluates sleep disordered breathing by measuring blood oxygen saturation and pulse with oximetry, airflow with a nasal cannula, snoring with a microphone, and head movement and position with an accelerometer.[7] However, the acceptability and success of self-application are not well established for such complicated home measurement devices.

To summarize, all of the devices that require measuring electrical activity in the brain will have a relatively high respondent burden and will likely be most useful in studies with primary hypotheses related to OSA or percentages of time in different sleep stages. This may be most feasible for modules of larger studies with strata defined for oversampling based on less burdensome sleep measures.

ACTIGRAPHY

Wrist actigraphy has increasingly been added to a number of large epidemiologic studies to estimate sleep characteristics (e.g., Van den berg et al 2009, Appelhans et al 2013, Lauderdale et al 2006, Fung et al 2013, Lauderdale et al 2014, Spira et al 2012, Anderson et al 2014).[8-14] Actigraphy does not directly measure sleep; estimates are based on arm movement. There are many different manufacturers and models, which is a challenge for comparisons across studies, and most researchers use the algorithms supplied by manufacturers. It is difficult to fully assess differences in how metrics are calculated between devices: although most manufacturers of research devices do share descriptive information about how their metrics are calculated, the programs themselves are not open. However, research grade devices do generally provide some less-processed data to the researcher, so it is possible to develop new metrics or validate the approach described by the manufacturer.

A typical actigraph is roughly the size of a digital watch; wearing an actigraph does not seem to alter sleep in that there is no "first-night effect" in studies with multiple nights,[15] making them a feasible objective way to assess sleep behavior. Actigraphs contain accelerometers that record activity counts in one or three directions and record them over brief epochs (generally 15 seconds to one minute). The level of activity in an epoch and the pattern of activity counts across contiguous epochs are used to infer sleep and wake in each epoch. They may be worn for days (models have varying capacity), storing the data that will subsequently be downloaded for analysis. Actigraphy data can be used to estimate a number of sleep characteristics, some of which appear to have been developed to offer rough equivalents to PSG metrics and some of which are unique to actigraphy. The raw or minimally processed data are available to create new metrics. Typical metrics are the duration of the interval from when an individual starts trying to fall asleep to when they finally awaken in the morning (aka rest interval), the interval between when someone actually falls asleep to when they finally awaken the next morning (aka sleep interval), and the difference between when they start trying to fall asleep and when they fall asleep (sleep latency), the sum of sleep epochs in the sleep interval (total sleep time), the sum of awake epochs in the sleep interval (wake after sleep onset – WASO), ratios of total sleep time to both the rest interval and sleep interval (aka sleep percent and sleep efficiency), various metrics that capture disruptions in sleep (sleep fragmentation) and the number of wake bouts during the sleep interval. The actual clock time of sleep onset and offset are also available.

Typically, studies collect 3 to 10 nights of contiguous recording, with 7 nights being most desirable for capturing weekday/weekend sleep variation. However, in the sleep sub-study for the National Social Life Health and Aging Study (NSHAP), there was no significant variation by day of the week for adults ages 60-90.[16] The day of the week effect on sleep may be more marked for younger adults partly because of higher proportion employed and partly because of aging-related declines in the ability to extend sleep on a weekend. Not only the mean of several nights, but also variability in sleep features or timing across nights may be a salient sleep characteristic. Variability can be minimally assessed with three nights of recordings, though four or more is preferable.

Actigraphs do not yield data on OSA or sleep stages. Clearly, arm motion is not the same as sleep, and there has been little research about relative accuracy for individuals with different health or other characteristics, or aging related changes in accuracy. How still someone is while awake in bed and how much someone moves while asleep surely vary between individuals and could be a source of differential accuracy of actigraphy sleep estimates. While the actigraphs used for research are relatively expensive, there is, in effect, no limit on how many times they can be reused (with fresh batteries). In general, total sleep duration correlates highly between PSG and actigraphy when both are measured concurrently. As aforementioned, the correlation is around 90 percent[15] with the previously described potential limitation.

CONSUMER TRACKING DEVICES

Consumer devices are popular, and it is not surprising that there has been a recent proliferation of studies exploring various devices' accuracy for sleep (and physical activity).[17-22] These studies, mostly conducted with small numbers of younger adults or adolescents, have compared various consumer devices to each other, to research actigraphs, or to PSG. Results are not consistent across devices, or even for the same device between different studies with regard to accuracy, sensitivity, or specificity in identifying wake and sleep epochs. With some important differences, consumer devices are generally similar to research actigraphy in that they estimate sleep characteristics using accelerometers. As currently available, they require a sophisticated user who can interface with a computer or smart phone. The wearer receives daily feedback, which makes wearing the device an intervention. The software is not open, so it is not really possible to know how the manufactures of the devices are calculating the sleep metrics. While they have great potential in research, it is not yet clear the extent to which they would be used for intervention versus data collection characterizing routine sleep behavior and experience. If they are an effective short-term intervention, they are less useful for non-intervention research without modification. It seems likely that modifications could be made to enable data submission to a research team without revealing it to the wearer on a daily basis. At the end of the study period, the data could then be shared with the respondent.

SURVEY QUESTIONS

Questions about sleep have face validity: they look like they measure what they purport to measure. However, there is convincing evidence that responses are not very accurate and are probably biased. There may be better ways to ask about sleep than those currently in use, but there has been relatively little methodological work about how best to ask about sleep.

DURATION

There are standardized questions or cascades of questions for asking about some health-related behaviors like smoking, but not for sleep. Many surveys include one question about sleep duration, but there are multiple ways of asking it. This variation

has been largely ignored in sleep research of all types, and studies are compared where the questions differ in ways that could be meaningful. Very little is known about the psychometric properties of different ways of asking about sleep, and few have noticed that heterogeneity in sleep duration correlates across studies might in part be related to differences in sleep questions.

Examples of duration questions from major surveys/studies:

- The question in the PSQI is: “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.)” and gives a response space labeled “HOURS OF SLEEP PER NIGHT.”
- The 1960 Cancer Prevention I cohort asked: “How many hours of sleep do you usually get at night?”
- The Nurses’ Health Study asked “How many hours of actual sleep do you get in a 24-hour period?”
- NHANES I Epidemiologic Follow-Up asked: “How many hours of sleep do you usually get at night (or when you usually sleep)?”
- The National Health Interview Survey asks: “On average, how many hours of sleep do you get in a 24-hour period?”
- The National Sleep Foundation Sleep in America Polls as two questions: “On workdays or weekdays, how many hours, not including naps, do you usually sleep during one night?” and repeats the question for “weekend nights.”

It is apparent that some of these questions actually ask for different quantities. There is nighttime versus 24-hour sleep, “usual” (modal) versus average, recent experience versus longer term. Some of these questions also introduce wording that may lead to an informational social influence[23] by implying weekday and weekend sleep durations should differ—that “actual” sleep differs from some other quantity that more readily comes to mind, or that some sleep occurs outside the nighttime.

Asking how long a respondent usually sleeps seems like a simple question. The format of it is similar to numerous survey questions that call on the respondent to retrieve readily accessible quantitative information: “What is your birth year?” or “How tall are you?” But how long one usually sleeps is not a number the great majority of individuals actually know. To answer this question accurately, one must first realize that this is a math problem and then determine an average time at which one falls asleep, an average time at which one wakes up, and then, holding those numbers in one’s head, carry out a mental subtraction of bedtime from wake time, which most often involves subtracting around midnight. In addition, for those who think they spend time awake during the night, they should estimate that duration and subtract it. This is a difficult and deliberative process that not all respondents will bother to undertake, instead producing a plausible answer quickly. (This perspective on the underlying problem with sleep duration questions draws on distinctions in intuition and accessibility described in many publications by Kahneman (e.g., Kahneman 2002[24]). Even if some make the calculation, the process has many points at which inaccuracies may be introduced.

Furthermore, the response format offered—whether or not there are fixed response categories of whole numbers or spaces labelled minutes or open spaces—might influence responses. To illustrate the complexity of figuring out an accurate response to a question about usual sleep hours, the Appendix includes part of the transcript from a cognitive interview we recently collected testing questions about sleep and physical activity. The section included is part of the transcript around sleep duration, collected as part of R01 AG051175. This was the first interview transcribed and was not selected to demonstrate a particularly complicated set of responses. (See Appendix)

To understand how sensitive sleep duration responses may be to question format, Lauderdale carried out a simple randomized survey experiment using an internet panel in which adults were randomized to one type of question variation: a single question about usual sleep hours versus two questions that separately asked about weekday and weekend sleep, but which otherwise had identical wording.[25] Mean sleep duration was significantly shorter for the single question than a weighted average (5/7 weekday + 2/7 weekend) from two questions, and the race/ethnicity and employment contrasts were different by question format, with the two-question format yielding the expected significant duration advantages for whites versus blacks and for the unemployed versus employed, while the single question did not. Everyone was also asked a logically related construct after the sleep hour question(s): how much sleep they needed to feel rested. The correlations between sleep duration and sleep need were significantly higher for the two-question format. This all suggested that the two questions—perhaps by virtue of the added weekday/weekend information inducing a more reasoned response rather than an intuitive answer—yielded more plausible responses. Regardless of why the answers differed, these results demonstrate that question wording can affect reported sleep durations in complicated ways and could lead to important differences between studies, such as the presence or absence of race disparities.

Alternate ways of asking about sleep duration in a survey format might avoid some of the limitations of this type of question. To avoid individuals resorting to easy answers when they do not know average sleep duration, one could ask—as a few surveys do—about usual bedtime and usual wake time and then let the investigators do the subtraction. Or, one could ask the respondent to keep a sleep log for one or more days. A single night's sleep log would introduce random noise due to night-to-night variation but may greatly reduce systematic bias introduced by asking people about "usual" or "average" quantities. Or a single night could be collected from 24-hour recall, as is done in time use studies. For employed individuals, seven days would be best, to capture any systematic differences between work days and non-work days.

A few studies among older adults have investigated how self-reported usual sleep duration compares to duration from several nights of actigraphy, or one night of PSG. In general, self-reported sleep is longer than measured sleep and correlates with actigraphy in the moderate range. In the Sleep Heart Health Study, one night of home PSG was compared to the next morning's estimate of that night and to previously reported usual sleep.[26] The single night correlation was only 0.16. Habitual sleep averaged an hour longer than PSG and the correlation was 0.18. Greater divergence

was associated with less education. The Rotterdam Study of older adults included several nights of actigraphy (average duration 6.5 hours), and subjects also estimated sleep each morning.[27] The morning estimates were 23 minutes longer than actigraph sleep, and about one-third had more than an hour difference. On average, women reported a quarter hour less sleep than men while their actigraph sleep was a quarter hour longer. An ancillary study to the Women's Health Initiative oversampled women reporting short or long sleep in response to "About how many hours of sleep did you get on a typical night during the past 4 weeks?" [28] Actigraph sleep averaged six hours, about 48 minutes shorter than self-reported sleep. The correlation was 0.48. Examining the actigraphy rest interval rather than total sleep time did not increase the correlation. An ancillary study to CARDIA compared 3 nights of actigraph sleep duration (average six hours) with self-reported measures in a middle-aged population.[29] The correlation was 0.47, with self-reported sleep averaging 48 minutes longer. Persons with mortality risk factors (more depressive symptoms, fair/poor self-rated health, or obesity) reported shorter sleep *at the same level of measured sleep*, suggesting that there may be a reporting bias that exaggerates the associations between worse health and shorter reported sleep. In the NSHAP ancillary sleep study, fair/poor self-rated health had different associations with self-reported sleep duration and actigraph sleep duration.[16]

Sleep Quality

As previously described, one common approach to determining sleep quality in a survey context is the PSQI. However, its use has largely been confined to smaller studies primarily about sleep or carried out by psychologists.

For epidemiologists, sleep quality has often been queried by a set of four questions that ask about the frequency of key features of an insomnia diagnosis. The common wording is

- "How often do you feel really rested when you wake up in the morning?"
- "How often do you have trouble falling asleep?"
- "How often do you have trouble with waking up during the night?"
- "How often do you have trouble with waking up too early and not being able to fall asleep again?"

There are single question versions of the above. For example, the English Longitudinal Study of Aging asked in one wave, "In the last 30 days, how much difficulty have you had with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning?" with response choices around severity rather than frequency: "None, Mild, Moderate, Severe, Extreme." Using a single question to encapsulate sleep quality concerns did not result in fewer overall questions, however, since there were three additional vignettes for anchoring: e.g., "Carol takes about two hours to fall asleep every night. She wakes up once or twice a night feeling panicked and takes more than an hour to fall asleep again. In the last 30 days, how much difficulty did Carol have with sleeping, such as falling asleep, waking up frequently during the night or waking up too early in the morning?"

In the NSHAP study, responses to the four insomnia symptom questions were compared to logically related actigraphy estimates (averages from three nights), e.g., do individuals who report trouble falling asleep have longer latency and less total sleep time?[30] The responses to the first question about feeling rested was not correlated with any actigraphy-estimated sleep characteristics. Responses to the other three questions seemed to indicate general sleep quality and not the specific sleep characteristic referenced. Each of the other three questions had several significant correlations with the actigraphy metrics, but generally not with the characteristics that each question seemed to reference. In some cases, the associations were not in the expected directions (e.g., people who reported waking up too early did wake up earlier on average, but they also had more mean total sleep time).

More general questions are also used to assess sleep quality. The Study on Global Aging and Adult Health asks: “Please rate the quality of your sleep last night. Was it very good, good, moderate, poor or very poor?” and also “Please rate the quality of your sleep the night before last. Was it very good, good, moderate, poor or very poor?” The Behavioral Risk Factor Surveillance System asks, **“**During the past 30 days, for about how many days have you felt you did not get enough rest or sleep? (number of days).” Validation for these questions would be helpful.

Sleep Disorder Risk

Sleep medicine specialists use survey instruments as initial screeners for sleep disorders, and some of these instruments have been adopted by population researchers. Assessing sleep apnea is of particular interest because of the evidence that it may contribute to increased cardiovascular risk and is strongly predicted by obesity. Thus, it might be a pathway linking obesity and cardiovascular outcomes. Some studies simply ask about apnea among a list of conditions, “Has a doctor ever told you that you have _____?” However, there are many undiagnosed cases of OSA.

The Berlin Questionnaire is often used to identify those at high risk of OSA in clinical research. It is a combination of questions about risk factors (e.g., obesity), purported outcomes, (e.g., hypertension), and symptoms (e.g., tiredness, snoring, or gasping for breath while sleeping).[31] Snoring by itself is sometimes used as an OSA risk indicator. Daytime sleepiness and snoring are both non-specific signs of OSA. The Epworth Sleepiness Scale is the most widely used instrument to measure daytime sleepiness. It asks how likely the respondent is to fall asleep in different situations, such as reading, driving, sitting or having a conversation.[32] These instruments (Berlin and Epworth) have been validated against clinical diagnoses and tests, with moderately good performance.

SUMMARY MEASUREMENT ISSUES

Survey questions about sleep duration and quality have been deployed in population surveys with little attention to validation and psychometric properties of questions. Single questions about sleep duration are very commonly used, but seem to be biased in ways that could create spurious correlations with health. As such, it is important not to use those correlations as intuitive evidence that simple sleep questions are adequate

for assessing sleep. Even in a survey context, there are likely better approaches to measuring sleep duration: questions about bedtimes and waking times in general or for a specific night or nights.

Sleep quality is an imprecise construct, but questions about quality may turn out to have greater validity than duration, when compared to objective sleep measures. There has been relatively little work around this, perhaps because insomnia, the most common complaint, is diagnosed based on history rather than objective measures. Older adults frequently complain about problems with insomnia symptoms, suggesting that the constraint on how much older adults sleep is the ability to initiate and maintain sleep, not a failure to prioritize sleep in a busy life.

Sleep medications may also be a good indicator of sleep quality, or independently important for health. Methods for accurately collecting medication data have been developed for population surveys, including direct observations of medications taken.

OUTSTANDING QUESTIONS & RESEARCH PRIORITIES ABOUT SLEEP AND HEALTH FOR OLDER ADULTS

The questions about sleep that we would like to be able to answer with population studies of older adults are: What are the direction and extent of the associations between sleep and health? Which domains of sleep are most important? (Duration? Timing? Insomnia? Consolidation? Satisfaction with sleep? Variability? Medications? Naps?) Would interventions that either generally improve sleep or address specific domains of sleep benefit health? If so, what aspects of physical or mental health would benefit?

Despite many articles having been published with promising findings of associations between sleep duration or quality and health outcomes, more work is needed to know whether they are likely to be causal. The first problem is that the great majority of studies are cross-sectional, although longitudinal results are increasing as cohorts with sleep questions mature. The main impediments to drawing on current evidence without further research are due to (1) the complexity of measuring sleep and the multifaceted nature of sleep, and (2) the possibility that reported correlations represent, at least in part, reverse causality and unmeasured confounding.

One priority is to determine whether there are better questions for survey data collection about sleep; whether survey questions are adequately valid and unbiased to be recommended; or whether actigraphy is instead preferable whenever feasible. Currently, survey questions have largely been confined to duration and the insomnia symptom questions, or general sleep quality/difficulty with likert scale responses (sometimes with anchoring vignettes). Other aspects of sleep, that may be relatively easy to ask about and yield adequately accurate data, may be worth testing: sleep medications, bedtimes and waking times, getting out of bed during the night, naps or variability. It may be more feasible to develop interventions for some sleep characteristics than others. For testing different kinds of survey questions, there

remains an issue of there being no gold standard for routine sleep behavior: the imperfect estimates from actigraphy will generally need to be the non-gold standard. Some focused methodological research among older adults that tests different types of questions and compares them (for accuracy) to sleep logs and actigraphy and also compares them (for importance) to various measures of function and health are needed to inform larger studies.

Because so much of the published literature on sleep and health has relied on simple questions about sleep duration, it is important to confirm these findings with different or better survey data collection approaches or with actigraphy. Sleep duration may be much more important – or much less important – as a determinant of health, quality of life and longevity than the current consensus among sleep researchers.

There are also challenges in using actigraphy: device choice, duration of protocol, and summary metrics.
Device choice. Choosing an actigraph from which minimally-processed data can be retrieved should be helpful for ascertaining the comparability of different algorithms from different devices, even if the planned analyses use the manufacturers' algorithms and metrics. Currently, it would make sense for future data collections to select devices with the newer three dimensional technology, which may prove to yield better summary data, either for sleep at night or for activity during the day.

Protocal length. Studies that have already collected 7 to 10 nights of data should investigate what would have been the loss in information or precision in mean estimates if they had collected fewer days, and how many days are needed to characterize night-to-night variability in sleep patterns. Continuing to recommend seven days or more, based on patterns for younger adults, may be unnecessarily burdensome and expensive in studies of older adults. Because devices are re-used many times during data collection, shorter protocols reduce the number of devices needed in a cohort.
Metrics. Actigraphy yields a very large amount of data, both the minimally processed activity counts (which most researchers do not use) and the summary metrics from the manufacturers software. The plethora of metrics is a challenge for research because there is generally no theory to guide the choice of metrics to investigate. Analyses with a long list of correlated sleep metrics may seem like data dredging instead of hypothesis-motivated studies.

Besides measurement, the other important methodological problem is confounding and reverse causality, that is the problem that reports of sleep problems or of non-normative sleep durations (whether shorter or longer than the generally accepted standard of seven to eight hours) are themselves the result of mental or physical health conditions. If this selection bias issue is not addressed in the design of an analysis, then incorrect inferences will follow attributing health outcomes to sleep as an exposure. It is relatively recently that the extent of social, economic and demographic predictors of sleep duration and quality have been appreciated, and many studies of sleep and health have not adequately controlled for even standard potential confounders (e.g., education). A

fuller appreciation of the correlates and predictors of different dimensions of sleep would help researchers design studies that avoided selection bias.

Beyond studies to address these methodological issues, there are key questions that build on previous findings in the sleep research literature:

- Do laboratory findings about restricted sleep durations in controlled environments and short-term cognitive and physiological deficits translate to real life associations between unmanipulated variation in sleep and health outcomes?
- Will interventions that increase sleep duration or improve sleep quality change health outcomes?
- What interventions work? Are consumer devices useful as an intervention?
- Is night-to-night sleep variability a determinant or consequence of physical and mental health problems?

In summary, sleep is important to older adults, and complaints about sleep are common. Population studies that have included a question or a few questions about sleep often find significant correlations between responses and numerous social, economic, demographic, mental health and physical health variables. But we do not know yet to what extent these correlations are valid and causal (and not due to measurement issues, confounding or reverse causality); how best and most feasibly to measure sleep in the context of population studies of older adults; and how much sleep interventions would benefit different dimensions of physical and mental health. Several cohorts of adults have added sleep measurement with actigraphy in the past few years, and so data to address some of these questions has likely already been collected. The sheer volume of data generated by actigraphy is a challenge to know which analyses to prioritize in these studies.

References

1. Ancoli-Israel S, Roth T. Characteristics of insomnia in the United States: results of the 1991 National Sleep Foundation Survey. I. *Sleep.* 1999;22 Suppl 2:S347-353.
2. Chen JH, Waite L, Kurina LM, Thisted RA, McClintock M, Lauderdale DS. Insomnia symptoms and actigraph-estimated sleep characteristics in a nationally representative sample of older adults. *J Gerontol A Biol Sci Med Sci.* 2015;70(2):185-192.
3. Ohayon MM, Roth T. Prevalence of restless legs syndrome and periodic limb movement disorder in the general population. *J Psychosom Res.* 2002;53(1):547-554.
4. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 1989;28(2):193-213.
5. Shepard JW, Jr., Buysse DJ, Chesson AL, Jr., et al. History of the development of sleep medicine in the United States. *J Clin Sleep Med.* 2005;1(1):61-82.
6. Redline S, Sanders MH, Lind BK, et al. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group. *Sleep.* 1998;21(7):759-767.
7. Westbrook PR, Levendowski DJ, Cvetinovic M, et al. Description and validation of the apnea risk evaluation system: a novel method to diagnose sleep apnea-hypopnea in the home. *Chest.* 2005;128(4):2166-2175.
8. Van den Berg JF, Miedema HME, Tulen JHM, Hofman A, Neven AK, Tiemeier H. Sex Differences in Subjective and Actigraphic Sleep Measures: A Population-Based Study of Elderly Persons. *Sleep.* 2009;32(10):1367-1375.
9. Appelhans BM, Janssen I, Cursio JF, et al. Sleep duration and weight change in midlife women: the SWAN sleep study. *Obesity (Silver Spring).* 2013;21(1):77-84.
10. Lauderdale DS, Knutson KL, Yan LL, et al. Objectively measured sleep characteristics among early-middle-aged adults: the CARDIA study. *Am J Epidemiol.* 2006;164(1):5-16.
11. Fung MM, Peters K, Ancoli-Israel S, et al. Total sleep time and other sleep characteristics measured by actigraphy do not predict incident hypertension in a cohort of community-dwelling older men. *J Clin Sleep Med.* 2013;9(6):585-591.
12. Lauderdale DS, Philip Schumm L, Kurina LM, et al. Assessment of sleep in the National Social Life, Health, and Aging Project. *J Gerontol B Psychol Sci Soc Sci.* 2014;69 Suppl 2:S125-133.
13. Spira AP, Covinsky K, Rebok GW, et al. Poor sleep quality and functional decline in older women. *J Am Geriatr Soc.* 2012;60(6):1092-1098.
14. Anderson KN, Catt M, Collerton J, et al. Assessment of sleep and circadian rhythm disorders in the very old: the Newcastle 85+ Cohort Study. *Age Ageing.* 2014;43(1):57-63.
15. 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(3):342-392.

16. Lauderdale DS, Chen J-H, Kurina LM, Waite LJ, Thisted RA. Sleep duration and health among older adults: associations vary by how sleep is measured. *Journal of Epidemiology and Community Health.* 2016;70(4):361-366.
17. Lee J, Finkelstein J. Consumer sleep tracking devices: a critical review. *Stud Health Technol Inform.* 2015;210:458-460.
18. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int J Behav Nutr Phys Act.* 2015;12:42.
19. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a Commercial Accelerometer with Polysomnography and Actigraphy in Children and Adolescents. *Sleep.* 2015;38(8):1323-1330.
20. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. *Med Sci Sports Exerc.* 2016;48(3):457-465.
21. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act.* 2015;12(1):159.
22. de Zambotti M, Baker FC, Willoughby AR, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiol Behav.* 2016;158:143-149.
23. Deutsch M, Gerard HB. A study of normative and informational social influences upon individual judgement. *J Abnorm Psychol.* 1955;51(3):629-636.
24. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. *Am Psychol.* 2003;58(9):697-720.
25. Lauderdale DS. Survey Questions About Sleep Duration: Does Asking Separately About Weekdays and Weekends Matter? *Behavioral Sleep Medicine.* 2014;12(2):158-168.
26. Silva GE, Goodwin JL, Sherrill DL, et al. Relationship between reported and measured sleep times: the sleep heart health study (SHHS). *J Clin Sleep Med.* 2007;3(6):622-630.
27. Van Den Berg JF, Van Rooij FJ, Vos H, et al. Disagreement between subjective and actigraphic measures of sleep duration in a population-based study of elderly persons. *J Sleep Res.* 2008;17(3):295-302.
28. Kripke DF, Langer RD, Elliott JA, Klauber MR, Rex KM. Mortality related to actigraphic long and short sleep. *Sleep Med.* 2011;12(1):28-33.
29. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? *Epidemiology.* 2008;19(6):838-845.
30. Chen JH, Waite L, Kurina LM, Thisted RA, McClintock M, Lauderdale DS. Insomnia symptoms and actigraph-estimated sleep characteristics in a nationally representative sample of older adults. *J Gerontol A Biol Sci Med Sci*. 2015;70(2):185-92.
31. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 1999;131:485–91.

32. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep.* 1991;14(6):540-545.

Appendix

Partial transcript from a cognitive interview testing sleep and physical activity questions. Data collected as part of R01 AG051175

I: Okay. Now I have some questions about your sleep. What time do you usually fall asleep at night?

R: [pause, sigh] I [pause] have been falling asleep [clears throat] about uh 9 o'clock. I finish- I get home later, cook, by the time I get that all done, put away, and I sit down and I find I fall asleep when I sit down [laughs]

I: [Laugh]

R: And I wake, that's the 9, or so, or 9:30. And then uh sometimes I'll wake up and it says 10:30, quarter 11. And then I get ready to go to bed.

I: Okay.

R: And I will sleep well and by 2:30, I'm awake again.

I: Okay.

R: And, uh, at that point I find difficulty getting back to sleep again. So my sleep habits are irregular. And I did have a test one day, and they told me I had a *slight* case of sleep apthea [sic] and uh to solve it, uh, you just sleep on your sides, not on your back. And uh to train yourself, you tie a, um, sew in your pajama tops, up at the neck, uh, a tennis ball.

I: Mm [laughs] Okay.

R: [laughs]

I: So, what time do you usually wake up in the morning?

R: In the morning I have the alarm set for uh, about 6:15. And then I get up and take my thyroid and I'll turn the TV on for a little bit to get the weather and so on. And um, I really shouldn't do that cuz that's when I'm sleeping [laughs] go to sleep again [laughs]

I: Okay

R: So, uh, but uh, then I go and get up and get serious about the day, about 7 [pause] 7:15.

I: So 7:15 is when you feel that you *get up* for the day?

R: Yeah, I have to get up for the day, get going. And that's when that 20 minutes or so, uh, period of getting tough [pause] to get going.

I: When you feel fatigued.

R: Yuh.

I: Okay! In the last week, on average about how many hours of sleep did you get a night?

[long pause]

R: Um [pause] I suppose six.

I: Okay. And how did you come to that number?

R: Well, I put the time I fall asleep before I go to bed together with what I am sleeping when I actually am prepared to go *to* sleep. And I think it's, because lots of times I just lay there and I'm wide awake. Then I will sometimes just get up and read. That makes me kind of sleepy after a while. And then I go, and I lay down, and then I fall asleep.

I: Okay so then -

R: And it's in sections. But I would say, average, six.

I: All right. Um -

R: I don't anyone else has habits like that, but I suspect they do.

I: [small laugh] Ok, so, you, said that you, um, well, I'll ask this later. [pause] What would you say is the least amount of sleep you got, um, one night last week?

R: [pause] Four hours

I: Okay. And what is the most?

R: Seven.

I: And what is the most common?

R: The six.