Measuring Sleep Habits Without Using a Diary: The Sleep Timing Questionnaire


Clinical Neuroscience Research Center, Western Psychiatric Institute and Clinic, University of Pittsburgh Medical Center

Study Objectives: To develop a single-administration instrument yielding equivalent measures of sleep to those obtained from a formal (2-week) sleep diary.

Design & Setting: A single-administration Sleep Timing Questionnaire (STQ) is described (and reproduced in the Appendix). Test-retest reliability was examined in 40 subjects who were given the STQ on two occasions separated by less than 1 year. Convergent validity was measured both by comparing STQ-derived measures with objective measures derived from wrist actigraphy (n=23) and by comparing STQ-derived measures with other subjective measures derived from a detailed 2-week sleep diary in two nonoverlapping samples (n=101, 93). Correlations of STQ measures with age and morningness-eveningness (chronotype) were also examined.

Subjects: The analyses used sample sizes of 40, 23, 101, and 93 (both genders, overall age range 20y - 89y). Most subjects were healthy volunteers; some Study 4 subjects were patients (enrolled in research protocols).

Results: Test-retest reliability for the STQ was demonstrated for estimates of bedtime (r=0.705, p<0.001) and waketime (r=0.826, p<0.001). Convergent validity using wrist actigraphy was demonstrated by correlations of 0.592 (p<0.005) for bedtime, and of 0.769 (p<0.001) for waketime. Diary studies indicated STQ bedtime and waketime data to be highly correlated (at about 0.8) with those obtained from a formal 2-week sleep diary. The STQ also provided data on estimated sleep latency and wake after sleep onset (WASO), which correlated reliably (at about 0.7) with average nightly ratings of these variables from a 2-week sleep diary. Mean estimated values of sleep latency and WASO from the two instruments were within 1 minute of each other. STQ-derived bedtimes and waketimes correlated with both age and chronotype in the expected direction (older subjects earlier, morning types earlier).

Conclusion: The STQ may be a reliable valid measure of sleep timing that could provide a time-efficient alternative to traditional sleep diaries.

Key Words: circadian rhythms, sleep, human, diary, ratings


INTRODUCTION

THE TIMING OF A SLEEP EPISODE IN RELATION TO THE INDIVIDUAL’S CIRCADIAN SYSTEM IS ONE OF THE MOST IMPORTANT DETERMINANTS OF SLEEP’S DURATION AND ARCHITECTURE. Whereas in the early days of sleep recording many laboratories were content to adopt a “one size fits all” attitude to bedtimes and waketimes, using a standard 10 p.m. bedtime and 6 a.m. waketime, for example, irrespective of what actually took place in the subject’s everyday life, current best practice dictates that polysomnographic (PSG) sleep recordings be tailored to start and end at the patient’s habitual bedtime and waketime, respectively.1 There are many good reasons for adopting such a rule. One is that there is a “wake maintenance zone” or “forbidden zone” for sleep onset in the 2 or 3 hours before habitual sleep onset, and a 10 p.m. bedtime for someone normally retiring at midnight is likely to lead to artificially long sleep-onset latencies.2,3 Another is that habitual waketimes may vary widely from one person to another, as a function of whether they are a “morning lark” or a “night owl”.4 Thus, an early morning laboratory waketime would be much more onerous for a “night owl” than for a “morning lark.” Even within an individual, morningness-eveningness can change through the middle years of life, with such changes accounting for many of the age-related PSG effects that are observed.5 Thus, even in clinical sleep studies, the use of arbitrary externally imposed sleep hours can result in atypical results and the potential for mistaken diagnostic impressions for a specific patient. It therefore behooves the investigator or clinician to get an accurate estimate of the habitual timing and duration of the sleep episode with a finer resolution than can usually be gleaned from a simple single question.

As noted above, the simplest way to determine habitual bedtimes and waketimes is to ask the patient or subject as part of a clinical interview. A problem with this approach is that answers can often be vague and inaccurate, with subjects merely reporting times to the nearest hour and, perhaps, being overly optimistic in their estimates of when they finally get to bed on most nights.6 Also, the answer can very often be “it depends,” with times varying as a function of many different social and occupational constraints. A better way is to have the patient or subject fill out a sleep diary or log for a week or two before coming to the laboratory. Sleep diaries have been used in studies dating back to the 1960s and beyond.7 Most sleep researchers have a preferred sleep diary which can vary in complexity from a simple chart marking bedtime and waketimes8 to a more detailed diary with additional ratings and measures of unwanted wakefulness.9 As has been confirmed by wrist actigraphy,10 and by ambulatory electroencephalographic monitoring,11 quite reliable estimates of bedtime and waketime can be obtained from such diaries. There are two main problems with diaries: 1) they require a nightly commitment from patients or subjects for a week or two, which they may find burdensome, and 2) diaries require access to the patient or subject well before the first laboratory PSG night. To overcome such problems, we sought to develop a questionnaire alternative to the sleep diary, yielding many of the same measures associated with a diary, but in a single-administration questionnaire format. The resulting Sleep Timing Questionnaire (STQ) is the subject of this paper, which has the goals of introducing the instrument, presenting information regarding test-retest reliability, and testing the convergent validity of the STQ in relation to both objective and subjective (diary) measures.

Disclosure Statement
Supported by NASA, NIA, NIMH and GCRC grants.

Submitted for publication February 2002
Accepted for publication September 2002
Address correspondence to: Timothy H. Monk, Ph.D., D.Sc., WPIC Room E1123, 3811 O’Hara St, Pittsburgh PA 15213 USA, Ph: 412-624-2246, FAX: 412-624-2841, Email: monkth@msx.upmc.edu

SLEEP, Vol. 26, No. 2, 2003 208
METHODS

Development of the STQ

The STQ (see Appendix) follows standard practice in our laboratories in first defining “good night time” (GNT) and “good morning time” (GMT). These are used in preference to the terms “bedtime” and “waketime” because of the need to determine when the patient or subject actually began to try and sleep, rather than the time at which they first retired to bed, perhaps to interact with their bedpartner, read a book, or watch TV. In a similar vein, GMT allows the specification that this is the time the patient finally finished the night of sleep, rather than waking in the early morning to use the bathroom, attend to a child, or let a pet out, for example, but then returning to bed. Having thus defined GNT and GMT, the questionaire then invites the patient or subject to separate nights before school or work (“worknights”) from nights before a day off, such as at weekends (“restnights”), dealing with the two situations separately.

The aim is then to focus the patient on the specifics of the earliest and latest GNT and GMT that they normally have. This part of the questionnaire is less to derive actual data endpoints but more to focus the patient on thinking about their actual bedtimes and waketimes, to avoid them merely responding with a “boilerplate” response such as “I go to bed at 11.” This leads naturally to the key question of their habitual GNT and GMT under worknight and restnight conditions, which can then be used to estimate appropriate timing of the laboratory PSG night, for example, or to provide estimates to be used in tailoring research protocols to the individual, or in deriving endpoints for survey studies. When overall estimates are required, averages between usual weeknight and restnight GNT and GMT with a 5:2 weighting of the weeknight to restnight estimates can be undertaken to yield an overall score. This 5:2 weighting is derived from the fact that people typically go to bed and wake up later on Friday and Saturday nights than on the other 5 nights of the week, a commonsense finding recently confirmed empirically. Such overall scores can then be used to approximate the average Time In Bed (TIB) i.e., the time at night devoted to attempting sleep. Clearly such procedures, while appropriate for most day workers, would need to be modified to be appropriate for shift workers, particularly those on rotating schedules. It should be noted that the studies to be reported here were specifically concerned with those on a regular night-sleeping schedule.

In addition, the STQ asks for estimates of GNT and GMT variability for worknights and restnights, which then yield a rough metric of bedtime/waketime stability. Diary measures often ask for an approximate estimate of subjective sleep latency and subjective wake after sleep onset (WASO) on a nightly basis. Clearly, to ask the patient to provide a single estimate of each is inherently inaccurate, requiring an estimated average from nights that may be very different from each other. However, to provide an admittedly very coarse estimate of each, questions regarding sleep latency and WASO comprise the final two questions in the STQ (see Appendix).

Study 1 Measures of Test-Retest Reliability

By scanning our database, we were able to identify a total of 40 subjects who had been administered the STQ on two separate occasions separated by less than 1 year (mean 105 days, range of durations: 6 days to 349 days). In most cases this double administration resulted from the STQ being given both at initial recruitment and during the first evening in the laboratory. In no case were the responses to the first STQ administration available to the subject when the second STQ was being completed. All subjects were healthy volunteers for various sleep research protocols. Ages ranged from 20 years to 82 years (mean: 46.3y, sd: 20.5y); there were 18 men and 22 women. All contributing studies conformed to the University of Pittsburgh Institutional Review Board (IRB) procedures regarding informed consent and conformed to the Declaration of Helsinki. Subjects who completed the study were typically paid $50 per night of PSG. The hypotheses to be tested by this study were that when the two STQ administrations were compared, there would be significant positive correlations between 1) overall GNT estimates at Time One versus overall GNT estimates at Time Two, and 2) overall GMT estimates at Time One versus overall GMT estimates at Time Two.

Study 2 – Concurrent Validity Assessment by Comparing STQ Measures of Bed Timing With Wrist Actigraphic Measures

By scanning our database, we were able to identify a total of 23 healthy control subjects who had been administered the STQ and who had subsequently (typically a few weeks later) completed a field study of wrist actigraphy lasting at least 1 week. In all cases the device used was the Actiwatch® device marketed by Minimitter Corp. of Bend, Oregon. Our own in-house software was used to determine mean activity onset and offset for the week as a whole, recognizing that the subjects were on a standard nychthemeral schedule (i.e., sleeping at night and...
awake during the day) and were usually forbidden from napping. For each subject, the overall GMT and GNT from the STQ were compared to the mean times of activity onset and activity offset (respectively), as determined from the 1 week wrist actigraphy trace. All subjects were volunteers for various sleep research protocols. Ages ranged from 23 years to 76 years (mean age: 45.1 y, sd: 17.3 y), there were 9 men and 14 women. All contributing studies conformed to the University of Pittsburgh IRB procedures regarding informed consent and conformed to the Declaration of Helsinki. Subjects who completed the study were typically paid $25 per week of actigraphy. The hypotheses to be tested by this study were that there would be significant positive correlations between 1) overall GNT estimates from the STQ with habitual time of activity offset estimates from wrist actigraphy, and 2) overall GMT estimates from the STQ and habitual time of activity onset estimates from wrist actigraphy.

Study 3 – Convergent Validity Assessment by Comparing STQ Measures of Bed Timing With Simple 2-week Diary Measures

A convenience sample of 101 healthy people aged 20 years to 59 years (53f, 48m; mean age: 33.5 y, sd: 13.2 y) was recruited using word of mouth and fliers. None were shift workers, recently returned from a different time zone, pregnant, or parents of children under 2 years of age. The study involved completing several single-administration questionnaires, including the STQ, and Composite Scale of Morningness (CSM) instrument, which yielded a score indicating the degree to which the subject was a morning-type versus an evening-type. Subjects were also required to complete a 2-week abbreviated version of the Social Rhythm Metric (SRM-5). This version comprised just five items, rather than the 17 items required for the standard Social Rhythm Metric. Each evening the subject was required to note the time at which each of five events was done (get out of bed, first contact with another person, start work, dinner, go to bed). The present analysis will focus on the first and last items (get out of bed, go to bed) of the SRM-5 completed for each of 14 consecutive nights. It should be noted that the final item (go to bed) corresponded to the time the subject first climbed into bed and thus did not correspond exactly to GNT as defined by the STQ (which corresponded to the time at which sleep was first attempted). In order to make comparisons between sleep timings from the single administration (STQ) and 2-week diary (SRM-5) instruments, worknight and restnight estimates from the STQ were averaged in the weighted ratio 5:2 to yield an overall GNT estimate and an overall GMT estimate, each of which could be compared to average “go to bed” and “get out of bed” estimates from the SRM-5 diary. The study conformed to the University of Pittsburgh IRB procedures regarding informed consent and conformed to the Declaration of Helsinki. Subjects who completed the study were paid $25.

The hypotheses to be tested by this study were that: 1) sleep timings (overall GNT and GMT) assessed by the STQ in this study would be correlated with those assessed from the same subjects using bedtime and waketime estimates from their 2-week SRM-5 diaries; 2) morningness would be correlated with GMT and GNT scores from the STQ (i.e., morning larks would have earlier GMT and GNT); and 3) age would be correlated with GMT and GNT scores from the STQ, with older subjects going to bed and waking up earlier.

Study 4 – Convergent Validity Assessment Using STQ and 2-week Pittsburgh Sleep Diary Measures in a Diverse Sample of Research Subjects

For a 33-month period (from February 1999 to November 2001), many research subjects who passed through our laboratory for PSG sleep studies, and who had also completed a 2-week sleep diary [the Pittsburgh Sleep Diary - PghSD®] prior to the PSG, were additionally asked to complete the STQ on their first evening in the laboratory. Completion of the STQ was made without reference to the PghSD. A total of 93 subjects were included (33 men, 60 women; mean age: 55.4 y, sd 18.4; range: 20 y - 89 y). None had previously been involved in Study 2. Of the 93 subjects, 40 were healthy controls, 53 had diagnosed illnesses (depression: 18, insomnia: 15, other sleep disorder: 5, other: 3); and 12 were caregivers of Alzheimer disease patients or organ transplant patients. All studies conformed to the University of Pittsburgh IRB procedures regarding informed consent for the particular study in which they were enrolled and conformed to the Declaration of Helsinki. Subjects were usually paid $50 per night of PSG.

This study focused on a comparison of various measures obtained from the 2-week PghSD diary and the single-administration STQ questionnaire. As was done in Study 2, for the STQ, overall scores were obtained by averaging worknight and restnight estimates of GNT and GMT with a 5:2 weighting of worknights versus restnights. Overall STQ estimates for GNT and GMT were then correlated with average GNT and GMT from the 2-week diary (PghSD). Also computed were correlations between estimates of sleep latency and WASO from the two instruments. The hypotheses to be tested by this study were that: 1)
sleep timings (overall GNT and GMT) assessed by the STQ in this study would be correlated with those assessed from the same subjects using the 2-week sleep diary; 2) overall mean bedtime and waketime estimates from the two instruments would be similar; and 3) estimates of sleep latency and WASO from the two instruments would be similar and would be correlated.

RESULTS

Study 1 Measures of Test-Retest Reliability

The scattergrams of Time One versus Time Two estimates of overall GNT and GMT are illustrated in Figure 1. In both cases the correlation was positive and significant (GNT: r=0.705, p<0.001; GMT: r=0.826, p<0.001), with the fitted regression line accounting for better than half of the variance. Almost all of the unexplained variance could be attributed to two outlying subjects in the measure of GNT—when these were removed, the test-retest correlation for this measure rose to 0.918.

Study 2 – Convergent Validity of STQ Measures of Bed Timing With Wrist Actigraphic Measures

The scattergrams of GNT (STQ) versus mean time of activity offset (actigraphy), and GMT (STQ) versus mean time of activity onset (actigraphy) are illustrated in Figure 2. In both cases the correlation was positive and significant (GNT: r=0.592, p=0.003; GMT: r=0.769, p<0.001), with the fitted regression line accounting for between 35% and 59% of the variance.

Study 3 – Convergent Validity Assessment by Comparing STQ Measures of Bed Timing With Simple 2-week Diary Measures

For overall scores of both GNT and GMT from the STQ, significant positive correlations emerged with equivalent estimates of “go to bed” and “get out of bed” from the 2-week SRM-5 diary (rho= 0.838, p<0.001; rho=0.860, p<0.001, respectively). Even with the slightly different definitions of bed timing used by the two instruments (see above), significant positive correlations emerged with equivalent estimates of “go to bed” and “get out of bed” from the 2-week sleep diary (PghSD) from the 93 subjects are given in Table 1. GNT estimates within 1 minute, when the STQ and sleep diary instruments were compared. The mean, median, and sd of the difference scores (in minutes) were 21, 13, and 44 for GNT; 14, 9, and 49 for GMT; 1, 1, and 12 for sleep latency; and 1, 0, and 27 for WASO. Statistically significant positive correlations between STQ and diary measures occurred for both GNT (r=0.788, p<0.001) and GMT (r=0.814, p<0.001), accounting for better than 60% of the variance. For measures of sleep latency and WASO (where some missing data reduced the sample sizes to 92 and 89, respectively), the correlations were also good (sleep latency: rho= 0.855, p<0.001; WASO: rho= 0.838, p<0.001), accounting for about 70% of the variance.

DISCUSSION

Clearly, no single-administration questionnaire can ever fully match the quality and richness of information provided by a 2-week sleep diary. However, the present results do suggest that the STQ provides acceptable data in answering the question of at what time a given subject habitually takes his or her sleep. In terms of reliability (Study 1), the STQ performed acceptably well, with test-retest correlations of better than 0.7, even after an average delay of 105 days. Validity assessments also confirmed the usefulness of the STQ, both with regard to objective assessments of the sleep-wake cycle made using wrist actigraphy and with regard to subjective assessments using diary instruments. Thus, correlations between STQ and actigraphy-based estimates of bedtime and waketime (Study 2) averaged around 0.6 to 0.7; those between STQ and diary estimates of bedtime and waketime (Studies 3 and 4) averaged around 0.8.

A second research question is that of whether the STQ would be sufficiently precise to test hypotheses regarding the influence of interindividual difference variables such as age and morningness on habitual sleep timing. In Study 3, the correlations with morningness score were about 0.7 for both overall GMT and overall GNT, indicating that about 50% of the variance was being explained. The STQ also did quite well (about 0.5 - 0.6) in tracking changes in sleep timing related to age. This was despite the absence of subjects conventionally considered “old” in this sample of 101 (the age range was 20y - 59y).

A third research question is that of whether the final two items of the STQ, asking questions regarding habitual sleep latency and WASO would be related to the average of nightly estimates from the 2-week sleep diary (Study 4). In both cases, the STQ did well, showing correlations of about 0.8 and mean estimates that were within a few minutes of the diary estimates. This is a particularly strong validation of the STQ in that nightly ratings were being compared to a single-administration overall estimate. It remains true, of course, that these correlations were between two subjective ratings, and it is quite possible that less satisfactory correlations would be obtained when objective measures of sleep latency and WASO are obtained using PSG, for example. A future paper is planned to address that topic.
CONCLUSIONS

The STQ appears to be a useful single-administration questionnaire for accurately assessing the habitual timing of a person’s sleep. STQ measures of sleep timing showed good reliability and validity, correlating well with both wrist actigraph-based and 2-week diary-based measures. Correlations were also obtained with individual difference measures such as age and chronotype. STQ questions relating to sleep latency and WASO correlated well with equivalent nightly diary measures averaged over 2 weeks.

ACKNOWLEDGMENTS

Primary support for this work was provided by NASA grant NAG9-1036. Additional support was provided by NASA grant NAG9-1234, National Institute on Aging grants AG-13396, AG-15138, AG-00972 and AG-15136, NIH grant MH 24652, and GCRC grant RR 00056. Sincere thanks to the many staff and faculty whose research projects contributed data to Study 4.

REFERENCES


APPENDIX

SLEEP TIMING QUESTIONNAIRE (STQ)

Name________________________
ID#________________________
Date________________________

SLEEP TIMING QUESTIONNAIRE (STQ)

This questionnaire asks about when you normally sleep. We are interested in getting as accurate a picture as we can of the times when you normally go to bed and get up. Please think carefully before giving your answers and be as accurate and as specific as you can be. Please answer in terms of a recent “normal average week,” not one in which you traveled, vacationed or had family crises. Thanks.

Please think of GOOD NIGHT TIME as the time at which you are finally in bed and start your day. Before a work day or school day,
what is your earliest GOOD NIGHT TIME? __________ am/pm
Before a work day or school day,
what is your latest GOOD NIGHT TIME? __________ am/pm
How stable (i.e., similar each night) are your GOOD NIGHT TIMES on a night before a day off (e.g. a weekend)? (circle one) 0-15mins. 16-30mins. 31-45mins. 46-60mins. 61-75mins. 76-90mins. 91-105mins 106-120mins.
2-3hours 3-4hours over 4hours

On a night before a day off (e.g. a weekend),
what is your earliest GOOD NIGHT TIME? __________ am/pm
On a night before a day off (e.g. a weekend),
what is your latest GOOD NIGHT TIME? __________ am/pm
On a night before a day off (e.g. a weekend),
what is your usual GOOD NIGHT TIME? __________ am/pm

How stable (i.e., similar each night) are your GOOD NIGHT TIMES on a night before a day off (e.g. a weekend)? (circle one)

Please think of GOOD MORNING TIME as the time at which you finally get out of bed and start your day.

Before a work day or school day,
what is your earliest GOOD MORNING TIME? __________ am/pm
Before a work day or school day,
what is your latest GOOD MORNING TIME? __________ am/pm
Before a work day or school day,
what is your usual GOOD MORNING TIME? __________ am/pm

How stable (i.e., similar each night) are your GOOD MORNING TIMES on a day before a day off (e.g. a weekend)? (circle one)

Please think of GOOD MORNING TIME as the time at which you are finally in bed and trying to fall asleep.

On the night before a work day or school day,
what is your latest GOOD NIGHT TIME? __________ am/pm
On the night before a day off (e.g. a weekend),
what is your latest GOOD NIGHT TIME? __________ am/pm

How stable (i.e., similar each night) are your GOOD NIGHT TIMES before a work day or school day? (circle one)

0-15mins. 16-30mins. 31-45mins. 46-60mins.
61-75mins. 76-90mins. 91-105mins 106-120mins.
2-3hours 3-4hours over 4hours

On a night before a day off (e.g. a weekend),
what is your earliest GOOD NIGHT TIME? __________ am/pm
On a night before a day off (e.g. a weekend),
what is your latest GOOD NIGHT TIME? __________ am/pm
On a night before a day off (e.g. a weekend),
what is your usual GOOD NIGHT TIME? __________ am/pm

These questions are about how much sleep you lose to unwanted wakefulness:

On most nights, how long, on average does it take you to fall asleep after you start trying? __________ minutes

On most nights, how much sleep do you lose, on average, from waking up during the night (e.g. to go to the bathroom)? __________ minutes