Contrasting Stylized Questions of Sleep with Diary Measures from the American Time Use Survey

Robin Kaplan, Brandon Kopp, and Polly Phipps
Bureau of Labor Statistics
Office of Survey Methods Research (OSMR)
QDET2
November 12, 2016

*Opinions expressed are those of the authors and do not constitute policy of the Bureau of Labor Statistics.
Sleep Duration in the ATUS

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Weekday</th>
<th>Weekend</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 to 24 Years</td>
<td>9.31</td>
<td>10.06</td>
</tr>
<tr>
<td>25 to 34 Years</td>
<td>8.44</td>
<td>9.52</td>
</tr>
<tr>
<td>35 to 44 Years</td>
<td>8.02</td>
<td>9.33</td>
</tr>
<tr>
<td>45 to 54 years</td>
<td>8.23</td>
<td>9.37</td>
</tr>
<tr>
<td>55 to 64 Years</td>
<td>8.21</td>
<td>8.97</td>
</tr>
<tr>
<td>65 Years and over</td>
<td>8.95</td>
<td>9.12</td>
</tr>
<tr>
<td>Total</td>
<td>8.54</td>
<td>9.4</td>
</tr>
</tbody>
</table>

* Data are annual averages for 2014.
Diary Collection of Sleep Estimates

**Introduction:**
"Now I’d like to find out how you spent your time yesterday, [day of week], [date], from 4:00 in the morning until 4:00 a.m. this morning. I’ll be asking where you were and who else was with you. If an activity is too personal, there’s no need to mention it."

**Sample interview excerpt:**
- I: "What were you doing at 4 a.m.?
- R: "I was sleeping."
- I: "What time did you wake up?"
- R: "7:00."
- I: "Okay. And what did you do next?"
Stylized Questions

- **National Health Interview Survey**
  - "On average, how many hours of sleep do you get in a 24-hour period?"

- **Behavioral Risk Factor Surveillance System**
  - "On average, how many hours of sleep do you get in a 24-hour period?"

- **National Sleep Foundation**
  - "..., about how much actual sleep would you estimate you typically get on work nights or weeknights?"
Sleep Duration in Other Surveys

<table>
<thead>
<tr>
<th>Survey</th>
<th>Hours of Sleep</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATUS (2014)</td>
<td>8.7</td>
</tr>
<tr>
<td>NHIS (2014)</td>
<td>7.1</td>
</tr>
<tr>
<td>BRFSS (2014)</td>
<td>7.0</td>
</tr>
<tr>
<td>National Sleep Foundation (2015)</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Stylized Questions: 1.6 Hours
## Potential Reasons for the Sleep Gap

### Diary
- List of activities
- Broad sleep lexicon
- 30-minute rule
- Naps are captured

### Stylized
- Single activity
- Sleep not defined
- Single sleep episodes
- Naps may not be captured
Response Process

Comprehension → Retrieval → Judgment → Response

* Tourangeau, Rips, & Rasinski (2000)
Questionnaire Evaluation Methods

- Behavior coding
- Cognitive interviews
- Quantitative online study
- Validation study
Behavior Coding Study
Research Interests

- Opportunity to observe the diary collection of sleep episodes in the production setting
- Identify concepts and tasks that respondents and interviewers may struggle with in reporting and recording sleep
- Detect practices that could be associated with sleep measurement issues
Study Methods

- N=104 transcripts, 2008 ATUS respondents
  - ~ 3 interviews per interviewer, for the 36 ATUS interviewers on staff
  - Non-random
  - Study sample is demographically representative of full 2008 sample
    - 8.53 hours of sleep on average, closely matching ATUS estimates

- Behavior coding of I and R interactions
  - Main unit of analysis is each conversation around sleep
    - Conversational “turns” related to sleep activity
    - Type of sleep activity; sleep time/duration; respondent recall and qualifications; interviewer probing
  - Reliability coding of 12% of transcripts
  - All results met a minimum kappa 0.4+, range was from 0.45 to 1.0
Conversational “turns”

Sample interview excerpt:
- I: "What were you doing at 4 a.m.?"
- R: "I was sleeping."
- I: "What time did you wake up?"
- R: "7:00."
- I: "Okay. And what did you do next?"
Reporting and Recording of Sleep

- Average of 21 turns to record sleep activities in an interview
  - Accounted for 12% of diary interactions

- Sleep mentions are times during an interview when respondent reports a wake time, sleep time or sleep duration
  - Each mention of waking or going to sleep took 6.7 turns
  - Reporting and recording most complex for going to sleep
    - Average of 11 turns for going to sleep compared to 3.9 for waking
Interviewer Probes

“What time did you wake up?”
- Leading question wording 69% of the time
  - What time did you get up?
  - What time did you wake up yesterday morning?

“I went to bed”
- Respondents used this phrase in 73% of interviews
- Interviewers followed up with a probe 95% of the time
  - Probes taught in training were used 42% of the time
  - Most common non-training probe was “did you go to sleep immediately?”
Time Qualifiers

Use of “about,” “around,” “maybe”

- An indicator of respondent confidence in answers
- Average of 1.8 qualifiers
- Respondents were slightly more likely to qualify responses when reporting time of falling asleep (61%) compared to waking up (49%)
Recall Strategies

- Explicit recall strategy reported in 25% of 331 sleep mentions across all interviews

![Bar chart showing recall strategies]

- Alarm
- TV
- Clock view
- Routine
- Other
Behavior Coding Summary

- Interactions are complex and respondents are uncertain, especially for falling asleep time
  - Respondents may have difficulty recalling or estimating sleep
  - Recall strategies are used to formulate waking and sleep response

- Interviewer probing may contribute to underreporting of time spent awake at night and in morning, inflating sleep time

- But behavior coding study only suggestive
  - Indirect evidence of measurement error
  - No follow up questions or stylized questions for comparison
Cognitive Interview Study
Methodology

- Interviewed participants in the lab
- Asked both diary and stylized questions
  - Counterbalanced order
- Retrospective probes

* Tourangeau, Rips, & Rasinski (2000)
Demographics

- Recruited from our participant database
- $N = 29$
  - 11 male; 18 female
  - Mean age = 46 ($SD = 14.05$)
  - Age range of 21 to 69 years old
Comprehension

What does "sleep" mean to you?

<table>
<thead>
<tr>
<th>Sleep Definition</th>
<th>Diary Mean Hours of Sleep</th>
<th>Stylized Mean Hours of Sleep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow sleep definition (n=15)</td>
<td>7.25</td>
<td>5.93</td>
</tr>
<tr>
<td>Broad sleep definition (n=13)</td>
<td>8.18</td>
<td>6.93</td>
</tr>
</tbody>
</table>
Diary Recall

- Wake time is easy to recall
  - Alarm
  - Look at clock when waking up
  - Structured schedule

- Sleep time is difficult to recall
  - Sleep time varies
  - Don't check clock
  - Infer time from TV program
## Stylized Estimates

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Frequency</th>
<th>Mean hours of sleep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recalled directly</td>
<td>10</td>
<td>6.22</td>
</tr>
<tr>
<td>Rate retrieval</td>
<td>9</td>
<td>7.02</td>
</tr>
<tr>
<td>Rate and adjustment</td>
<td>4</td>
<td>5.62</td>
</tr>
<tr>
<td>Calculation</td>
<td>4</td>
<td>5.50</td>
</tr>
<tr>
<td>Estimate/Guess</td>
<td>2</td>
<td>6.50</td>
</tr>
</tbody>
</table>
What is the \(\text{maximum/minimum}\) number of hours appropriate to sleep per night?

<table>
<thead>
<tr>
<th></th>
<th>Diary</th>
<th>Stylized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum ((M = 8.87) hrs)</td>
<td>(r = 0.39^*)</td>
<td>(r = -0.10, \text{n.s.})</td>
</tr>
<tr>
<td>Minimum ((M = 5.97) hrs)</td>
<td>(r = 0.23, \text{n.s.})</td>
<td>(r = 0.40^*)</td>
</tr>
</tbody>
</table>

\(^*p < .05\)
Summary

- Comprehension
  - Broader definition = more sleep

- Recall
  - Easy to recall wake times
  - Hard to recall sleep times

- Estimation
  - Wide range of strategies that may be prone to measurement error

- Reporting
  - Social desirability concerns
Quantitative Study
Design

- Participants completed Diary and Stylized sleep measures
  - Embedded questions about other activities

- 2 Definition Conditions
  - With definitions; Without definitions

- 2 Order Conditions
  - Diary first; Stylized first
Participants

- 1233 participants completed the study
  - Recruited through Amazon Mechanical Turk
  - 46% male; 54% female
  - Mean age = 36 ($SD = 11$; range = 19 - 77)
Definition Condition

**Sleeping**: By sleep, we mean the number of hours you actually spend sleeping. This may be different from the number of hours you spend in your bed, time you spend preparing to go to sleep, resting with your eyes closed but not actually asleep. Please **include** any times you were sleeping during the day (or napping)

- Also read definitions for other activities (exercise, work)
Sleep Questions

- Modified version of the ATUS interview
- Past Week Stylized
  
  “Thinking back to the past week (that is, during the previous 7 days), how many hours of sleep did you get on average each weeknight (excluding weekends)?”
Results

Means and Standard Deviations

<table>
<thead>
<tr>
<th>Hours of Sleep</th>
<th>Diary</th>
<th>Past Week Stylized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.95</td>
<td>7.27</td>
</tr>
</tbody>
</table>

Response Distribution

- Diary
- Past Week Stylized

Hours of Sleep: 0 to 18
Results

2 (Question Type) X 2 (Definition) X 2 (Order) ANOVA

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>F Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question Type</td>
<td>1</td>
<td>200.49</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>Question Type X Definition</td>
<td>1</td>
<td>5.84</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Question Type X Order</td>
<td>1</td>
<td>29.49</td>
<td>p &lt; 0.01</td>
</tr>
<tr>
<td>Question Type X Definition X Order</td>
<td>1</td>
<td>0.18</td>
<td>p = 0.67</td>
</tr>
<tr>
<td>Residuals</td>
<td>1229</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Main effect of Question Type
- Question Type X Order Interaction
- Question Type X Definition Interaction
Results

**Question Type X Definition**

- Without Definitions (n = 622):
  - Diary: 8.0
  - Past Week Stylized: 7.2

- With Definitions (n = 611):
  - Diary: 7.9
  - Past Week Stylized: 7.3

**Question Type X Order**

- Diary First, Stylized Second (n = 626):
  - Diary: 8.4
  - Past Week Stylized: 7.5
  - Corr = 0.34

- Stylized First, Diary Second (n = 607):
  - Diary: 7.5
  - Past Week Stylized: 7.1
  - Corr = 0.54
Summary

- Diary > Stylized
- Definitions brought measures closer together
- Order effect
Validation Study
Wearable Activity Devices

Sleep

LAST NIGHT

Awake

Restless

4 hrs 21 min

2 min (tx)

29 min (19x)

THU

12PM 2

4

6

8

10

12AM 2

4

6

8

10

12PM

FRI

PAST 7 DAYS

WED

TUE

MON

SUN

SAT

FRI

12PM 2

4

6

8

10

12AM 2

4

6

8

10

12PM

SLEEP

RESTLESS

AWAKE
Methodology

2 visits one week apart

▲ Visit 1:
  - Demographic questions
  - Fitbit instructions

▲ Visit 2:
  - ATUS interview
  - Stylized questions
  - Debriefed about Fitbit-recorded sleep data
Participants

- 19 participants from the Washington, DC metro area
  - 13 female, 6 male
  - Mean age = 44.58 years
Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Hours of Sleep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diary</td>
<td>7.4</td>
</tr>
<tr>
<td>Stylized</td>
<td>6.4</td>
</tr>
<tr>
<td>FitBit</td>
<td>6.8</td>
</tr>
</tbody>
</table>
Debriefing

- 12/19 thought the Fitbit sleep data was about right
  - 5 thought it underestimated
  - 2 thought it overestimated

- Recall aid
  - "Oh yeah, I hit the snooze button that morning"
Logistics with Validation Research

- Measurement error
  - Records tossing and turning as times awake (underestimates)
  - Records inactivity as naps (overestimates)

- User error
  - Did not adhere to instructions to wear device each night (n=5)
  - Only wore device one work night of the week (n=7)
  - Lost device (n=2)
  - Device fell off night before Visit 2 (n=1)
Summary

- Preliminary evidence that an objective sleep measure might fall in between diary and stylized reports of sleep
  - But has its own set of measurement and user error
- Device data can be used to aid memory in time use research
- Conduct more interviews in the future
Conclusions
Multi-method approach

- **Behavior coding**
  - Pros: Uses actual production survey transcripts, examine questions, probes, answers, unscripted conversation
  - Cons: Time consuming!, Only indirect evidence of measurement error, no comparison to stylized questions

- **Cognitive interviews**
  - Pros: Insight into each stage of the response process, generate hypotheses
  - Cons: Small sample size, limited geographic region, qualitative data
Multi-method approach

**Quantitative study**
- Pros: Collect data quickly with large sample, make statistical comparisons
- Cons: Convenience sample (voluntary panel), no data on accuracy

**Validation study**
- Pros: self-report data validated against objective data, relatively inexpensive with new technologies
- Cons: logistically complex, user error, and measurement error
Multi-method approach

**Recommendations:**

- Weigh the pros and cons of each method
- Capitalize on the strengths of each type of method
- Build results off of one another to gain deeper understanding of an issue
Contact Information

Robin Kaplan, Brandon Kopp, and Polly Phipps
Bureau of Labor Statistics
Office of Survey Methods Research (OSMR)
QDET2
November 12, 2016

E-mail:
Kaplan.Robin@bls.gov
Phipps.Polly@bls.gov