

Using Online Testing and Wearable Devices to Pretest Diary and Stylized Sleep Measures

Robin Kaplan, Brandon Kopp, and Polly Phipps

Bureau of Labor Statistics

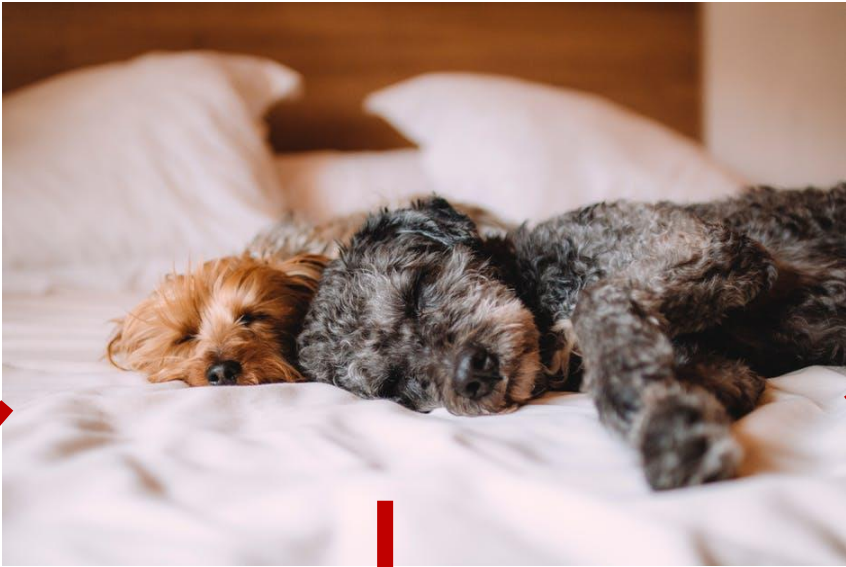
Office of Survey Methods Research (OSMR)

FedCASIC 2019

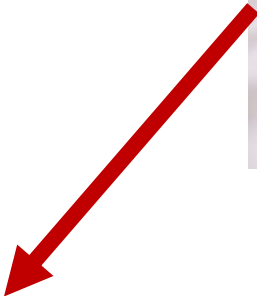
Importance of Sleep Measures



Diary



Stylized





American Time Use Survey (ATUS)

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?"

Self-Report Sleep Measures



Diary

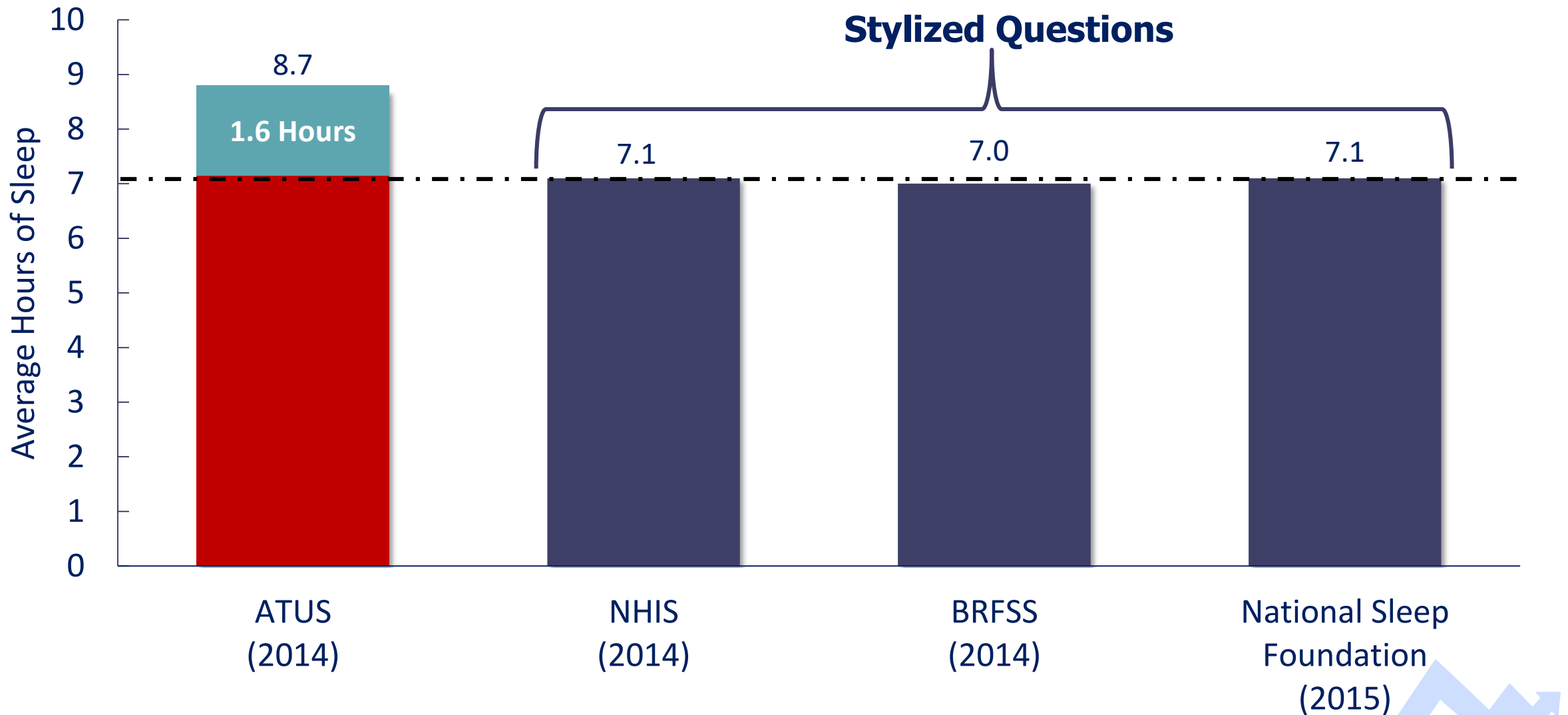
- Respondents report on all of their activities during a specified time (e.g., prior 24-hours), including sleep
- Most reliable, but expensive (Schulz & Grunow, 2012)



Stylized

- Respondents report on the average, typical, or usual amount of time spend sleeping
- Less reliable, but easy and cheap to administer (Bonke, 2005; Kan & Pudney, 2007)

Sleep Duration in ATUS vs. Other Surveys



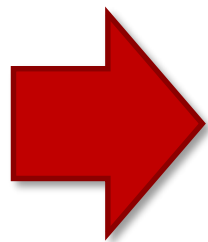
Cognitive Interview Study



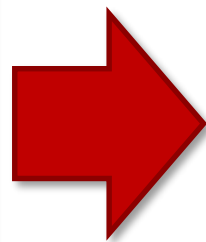
Methodology

- Interviewed 29 participants in the Washington, DC metro area
- Asked both **diary** and **stylized** questions about sleep
- Retrospective probes on response process at end of interview

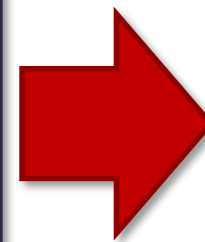
Comprehension



Retrieval



Judgment



Response

Tourangeau, Rips, & Rasinski (2000)

Participants

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

Results

■ Comprehension

- ▶ Broader definition = more sleep; Stylized question = continuous nighttime sleep

■ Recall

- ▶ Easy to recall wake times (alarm)
- ▶ Hard to recall sleep times (TV)



■ Judgment

- ▶ Wide range of strategies that are prone to measurement error (e.g., rate retrieval, rate and adjustment, calculation, guessing)

■ Reporting

- ▶ Context effects; easy to edit stylized reports; give "normative" amount

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)

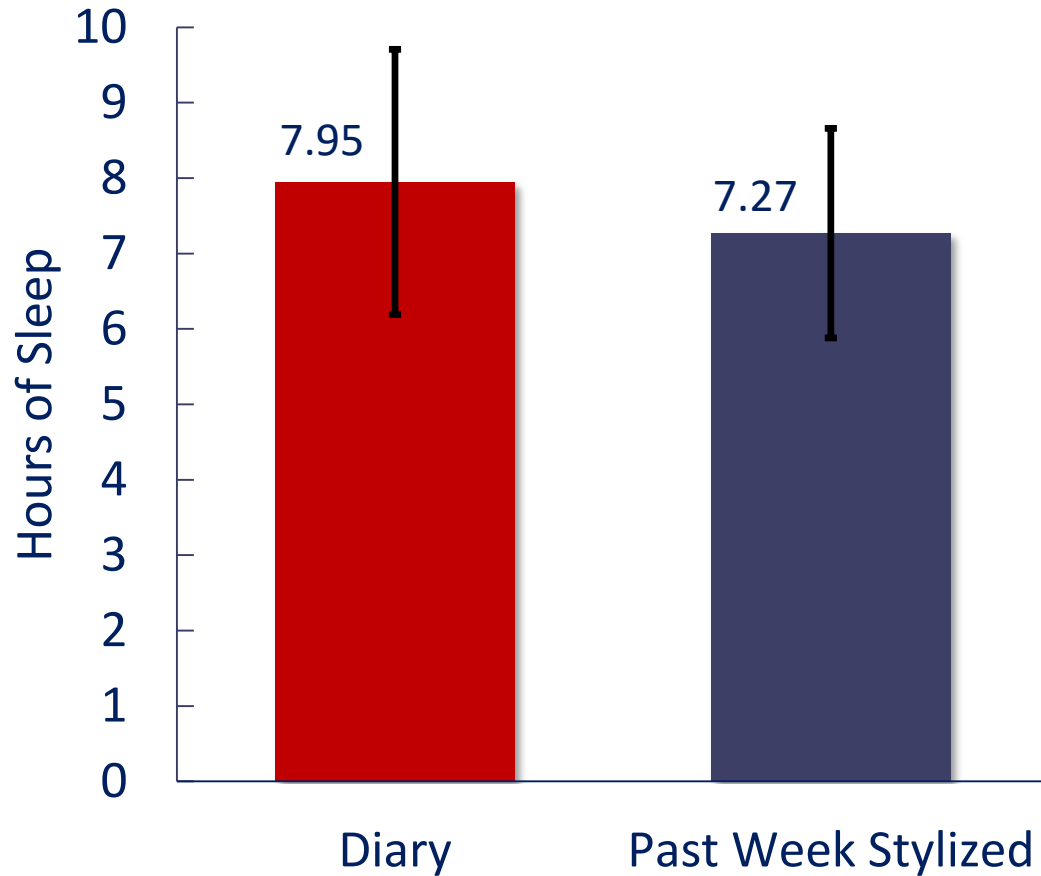
Sleep Questions

- 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)?”

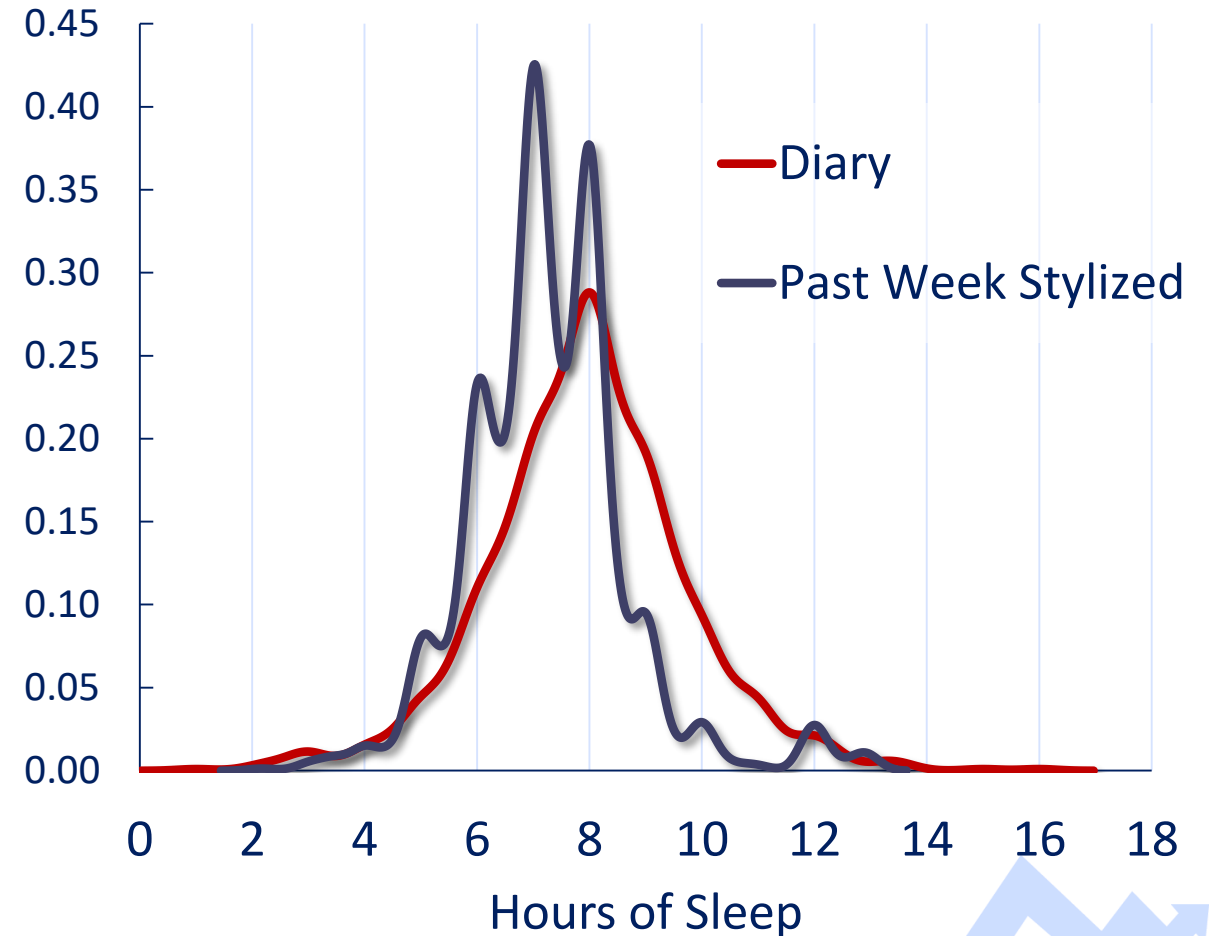
	Activity	Start Time (Hour)	Start Time (Minute)	Start Time (AM or PM)	End Time (Hour)	End Time (Minute)	End Time (AM or PM)
2	Grooming	7	:00	AM	7	:30	AM
3	Eating/Drinking	7	:30	AM	7	:45	AM
4	Travelling & going from place to place	7	:45	AM	8	:15	AM
5	Work & Work-Related activities	8	:15	AM	12	:45	PM

Results

Means and Standard Deviations



Response Distribution



Results

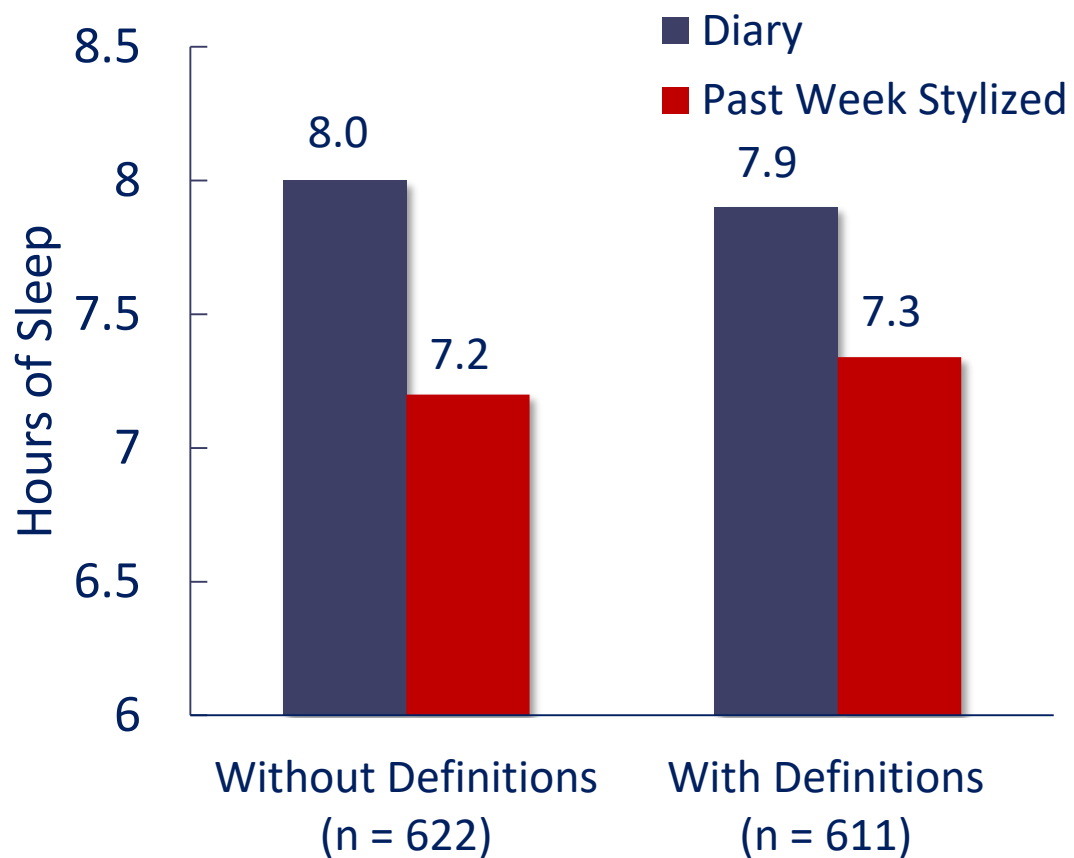
■ 2 (Question Type) X 2 (Definition) X 2 (Order) ANOVA

	df	F Value	p-value
Question Type	1	200.49	p < 0.01
Question Type X Definition	1	5.84	p < 0.05
Question Type X Order	1	29.49	p < 0.01
Question Type X Definition X Order	1	0.18	p = 0.67
Residuals	1229		

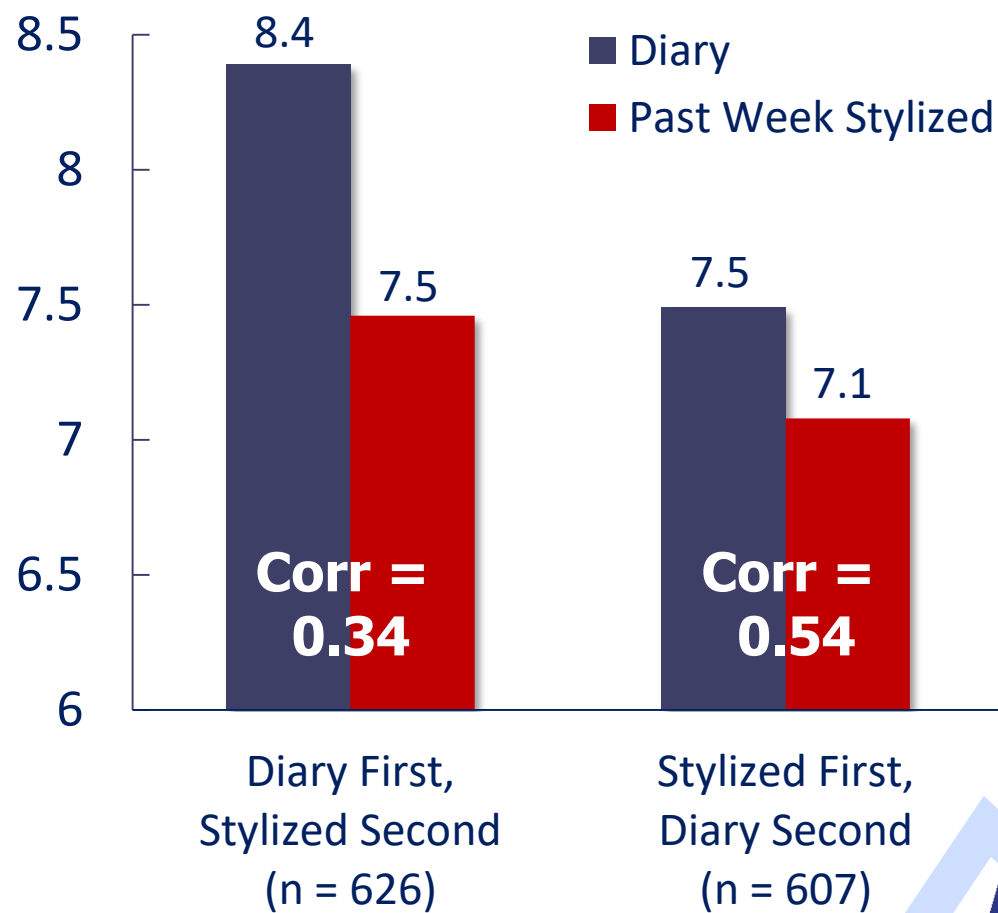
- Main effect of Question Type
- Question Type X Order Interaction
- Question Type X Definition Interaction

Results

Question Type X Definition



Question Type X Order



Summary

- Diary > Stylized
- Definitions brought measures closer together
- Order effect



Validation Study



FitBit Charge



Sleep

LAST NIGHT

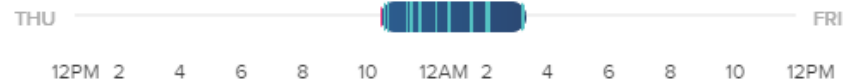
4 hrs 21 min

AWAKE

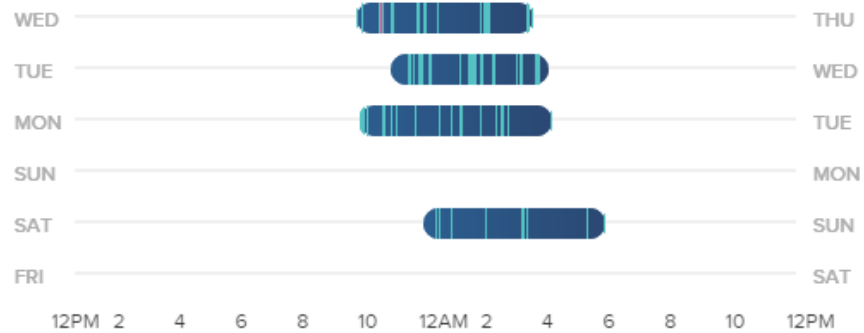
2 min (1x)

RESTLESS 

29 min (19x)



PAST 7 DAYS



■ SLEEP ■ RESTLESS ■ AWAKE



Methodology

■ Week 1 Visit

- ▶ Demographic questions
- ▶ FitBit instructions
- ▶ Wore device for 1 week



Visits were
always on
weekdays
(Tues-Fri)

■ Week 2 Visit

- ▶ Diary questions
- ▶ Stylized questions
 - General
 - Week
- ▶ FitBit data comparison



Participants

- 35 participants from the Washington, DC metro area
 - ▶ 22 female, 13 male
 - ▶ Mean age = 42.77 years old
 - ▶ Mean household size = 2.83

Sleep Duration Measures



Diary

- Total number of hours participant reported sleeping in prior 24-hour period
 - Weekdays only



Stylized

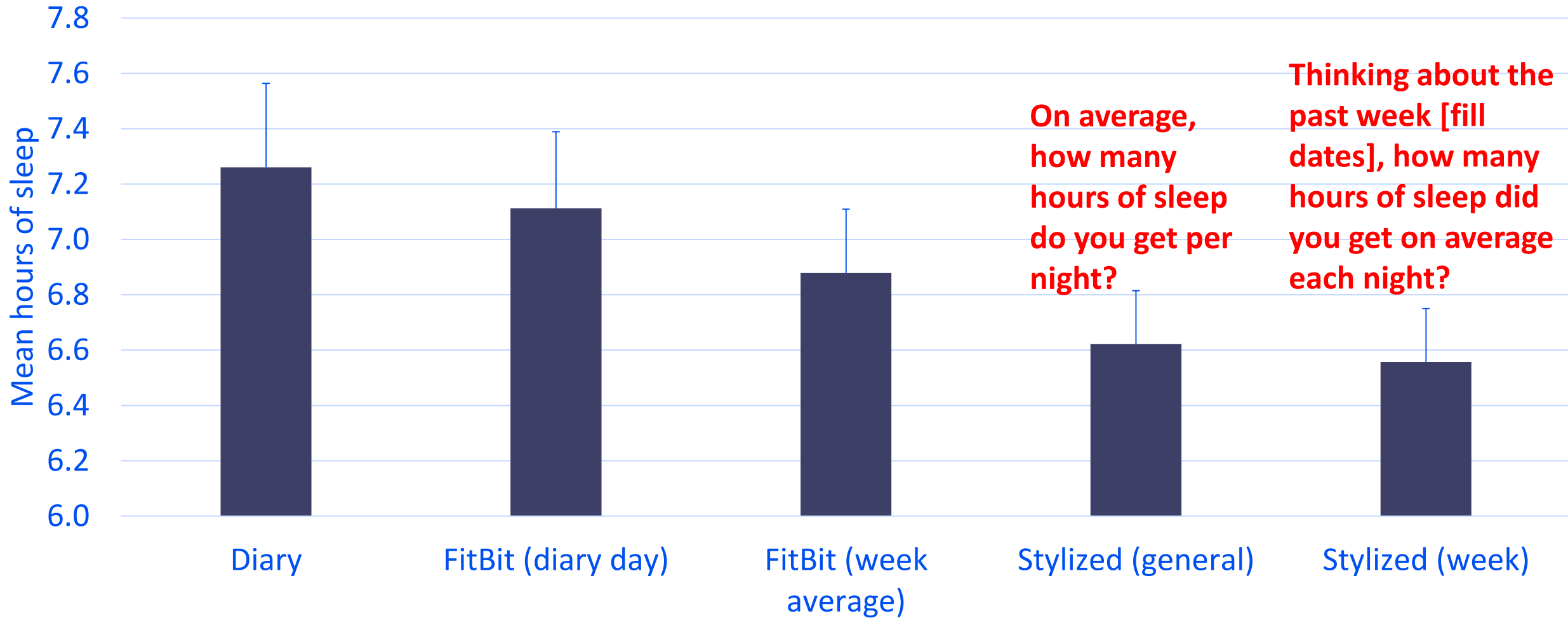
- General: On average, how many hours of sleep do you get per night?
- Week: Thinking about the past week [fill dates], how many hours of sleep did you get on average each night?
 - Weekdays
 - Weekends



FitBit

- Number of hours of FitBit-recorded sleep on:
 - Diary day
 - Overall week average
 - Weekdays
 - Weekends

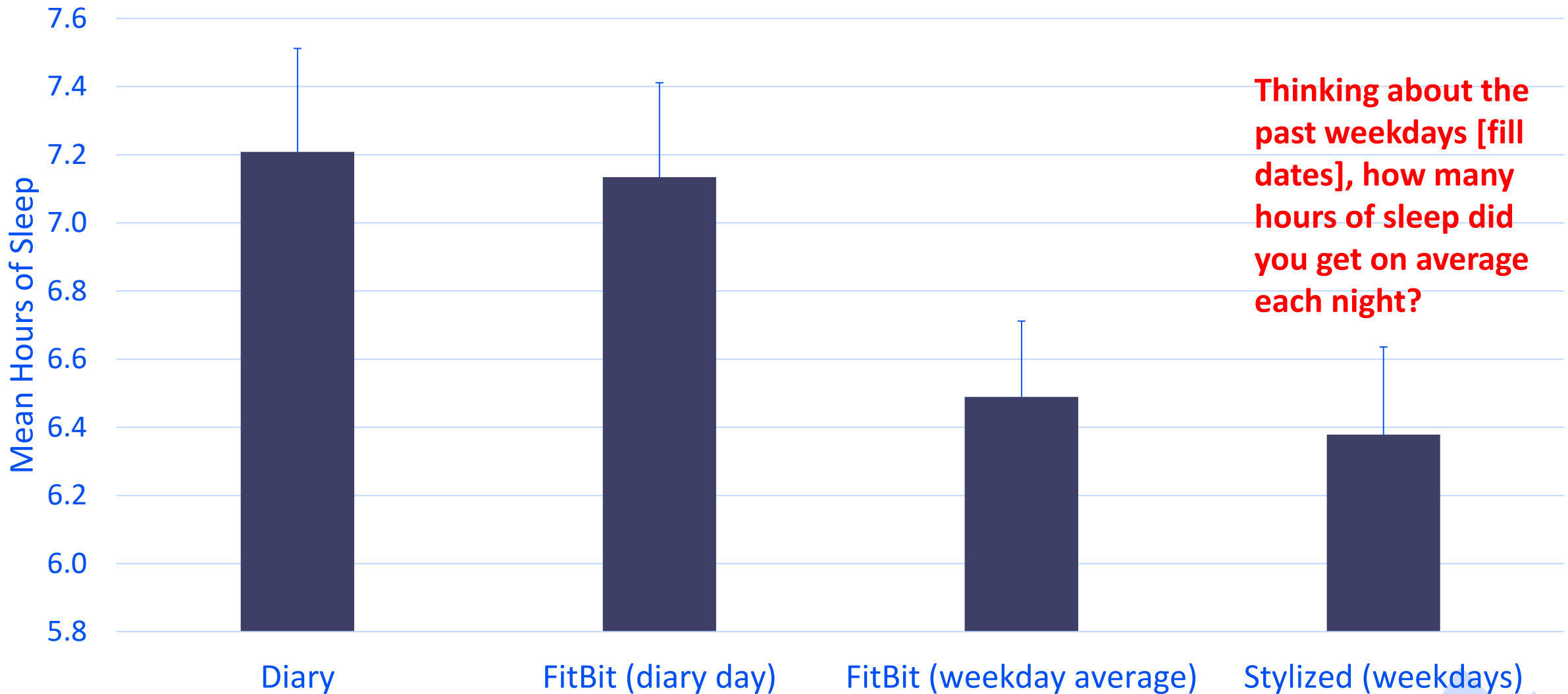
Sleep Duration Measures (Overall)



Diary & FitBit (diary day) exceeded both Stylized measures ($ps < .05$), $\eta^2 = .18$



Sleep Duration Measures (Weekdays)



Diary and FitBit (diary day) exceeded Stylized ($p < .005$), $\eta^2 = .14$



Agreement Across Sleep Duration Measures



Diary



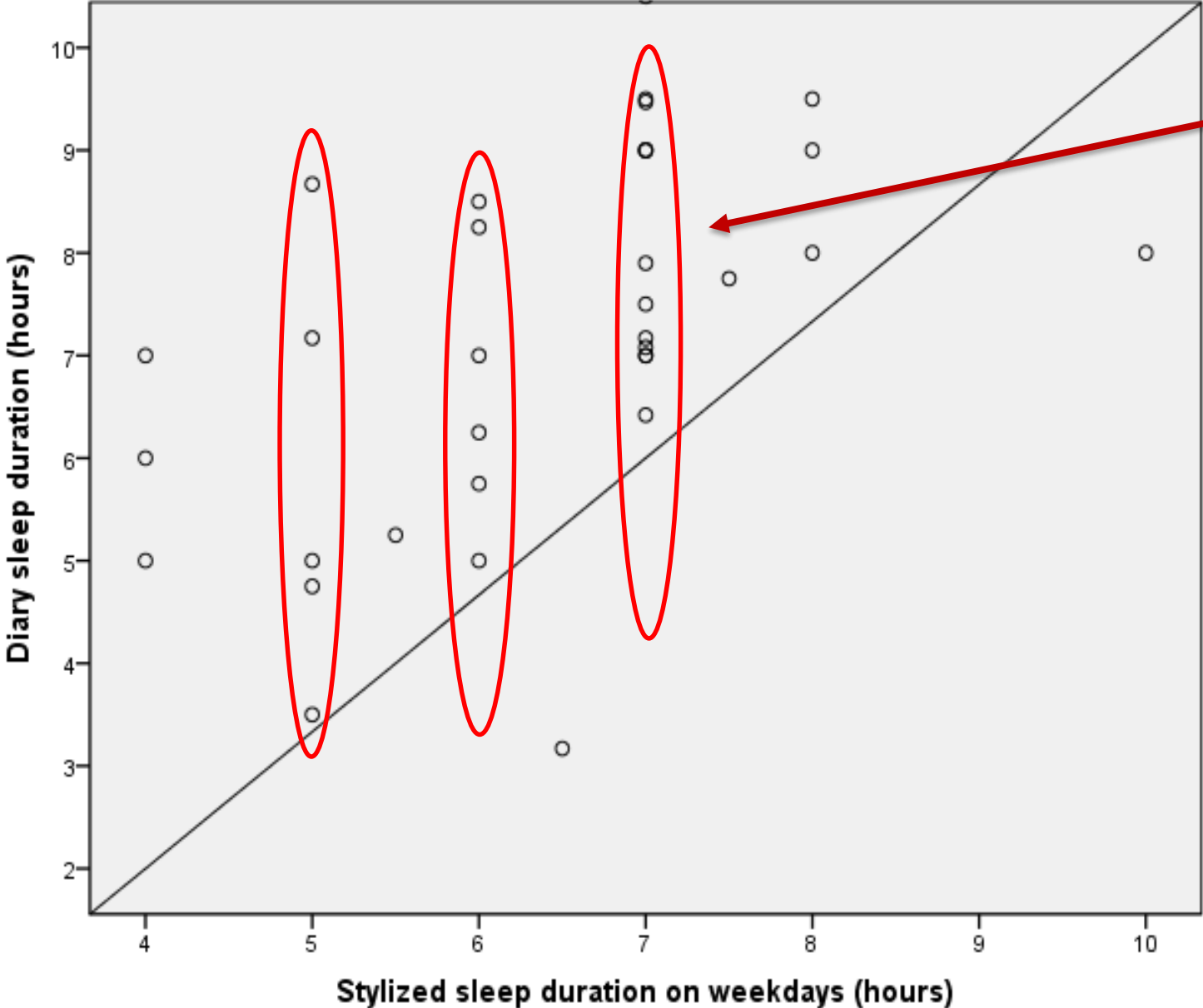
Stylized



Fitbit

Diary and Stylized Sleep Duration (weekdays)

Stylized Question:
Thinking about the **past weekdays** [fill dates] how many hours of sleep did you get on average each night?



Rounding in Stylized Measure

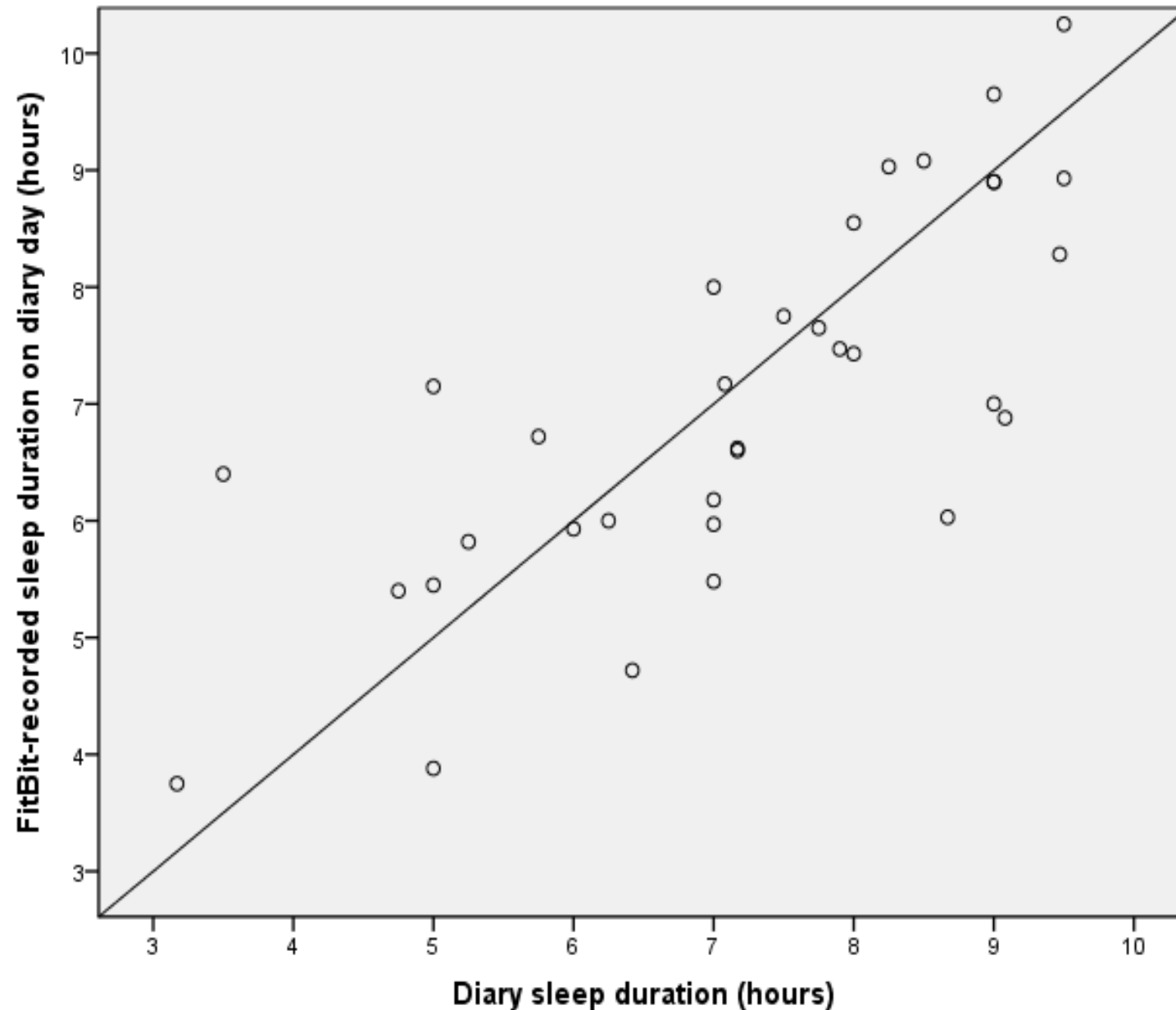
Fair agreement

ICC = .49
(CI .19-.71)
rho = 0.61

ps < .05



Diary and FitBit-Recorded Sleep (on diary day)



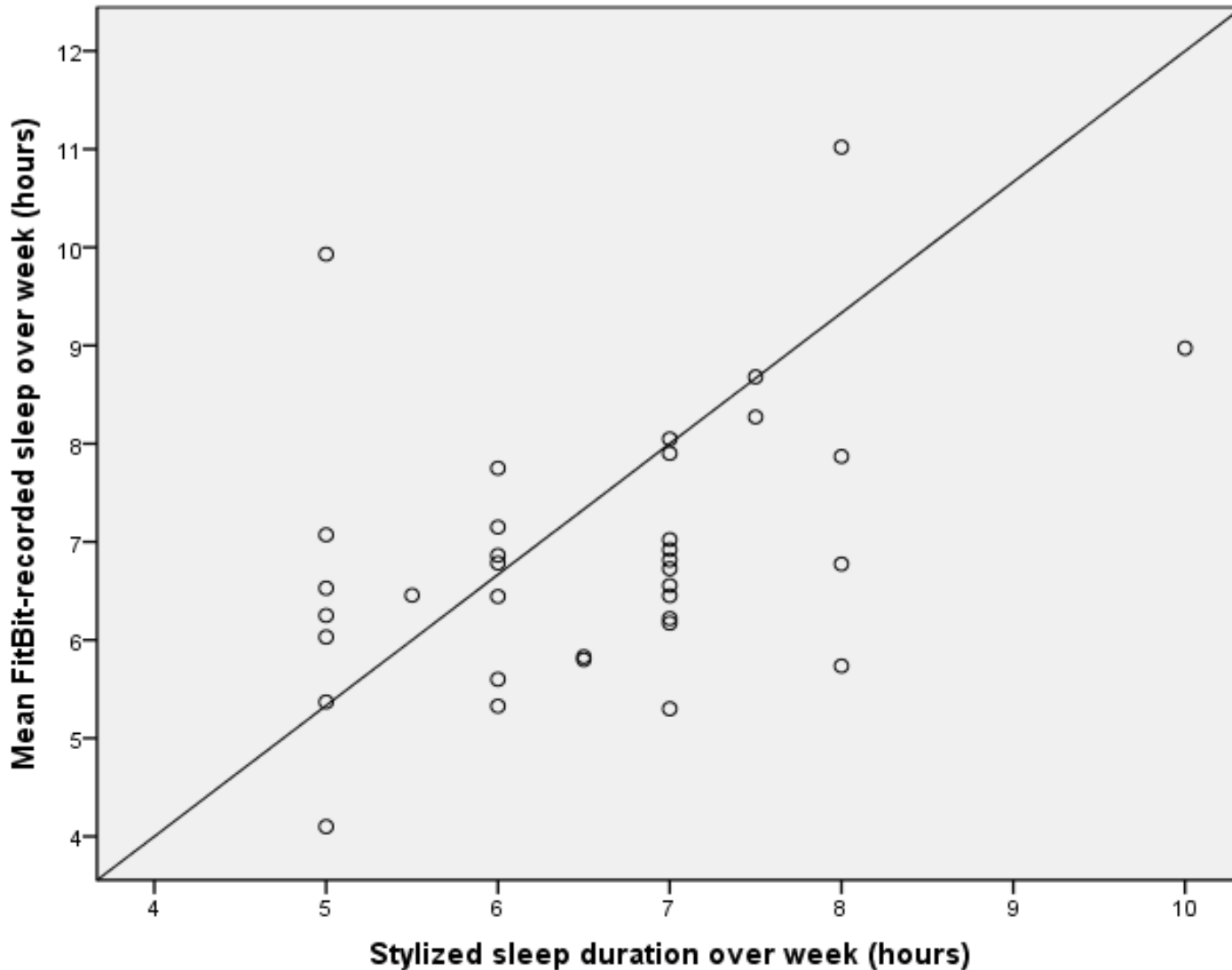
**Excellent
agreement**

ICC = .76
(CI .60-.88)
 $r = 0.78$

$ps < .001$

Stylized and FitBit-Recorded Sleep (over week)

Stylized Question:
Thinking about the **past week [fill dates]** how many hours of sleep did you get on average each night?



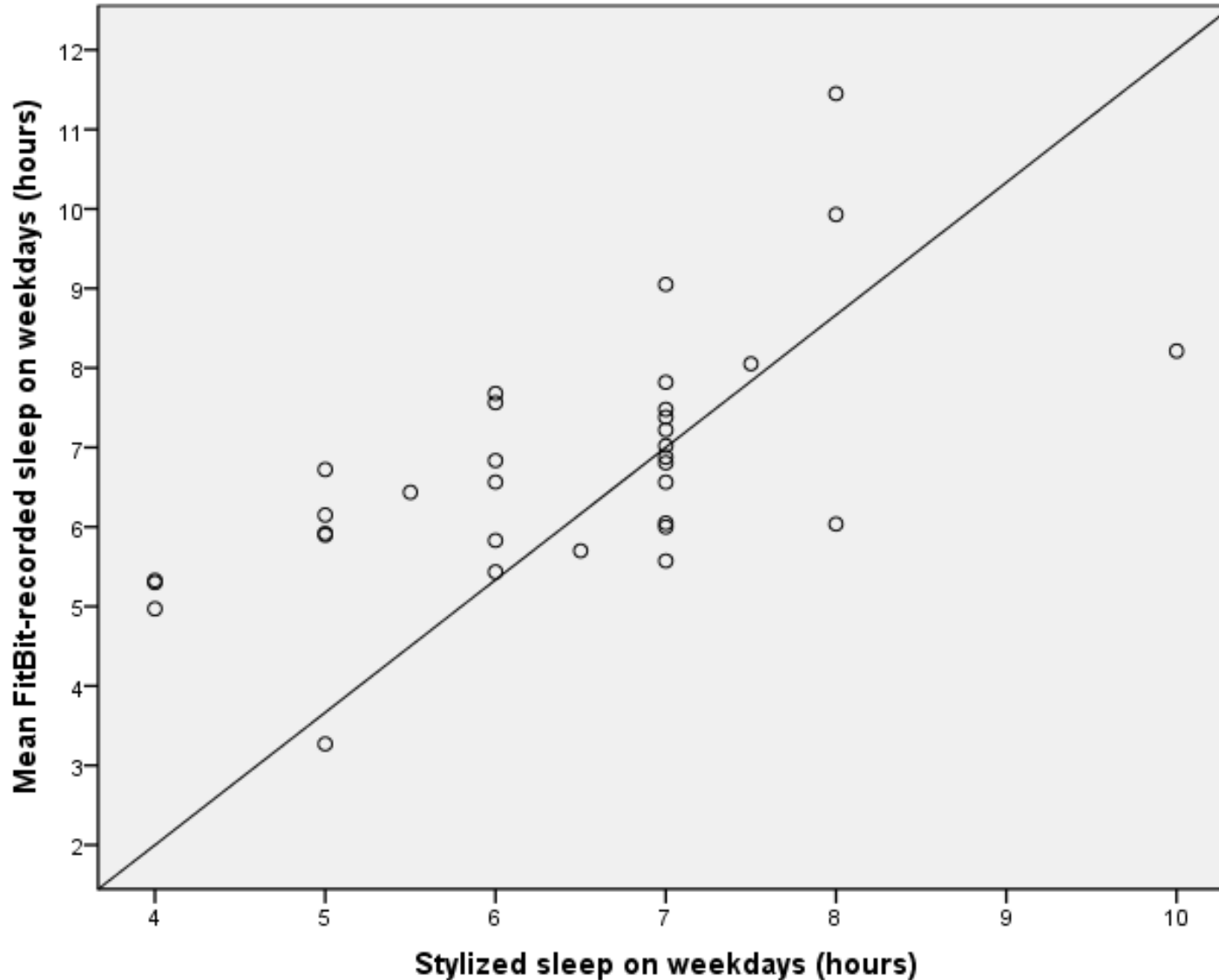
Fair agreement

ICC = .40
(CI .08-.64)
 $\rho = 0.36$

$p < .05$

Stylized and FitBit-Recorded Sleep (weekdays)

Stylized Question:
Thinking about the **past weekdays** [fill dates] how many hours of sleep did you get on average each night?



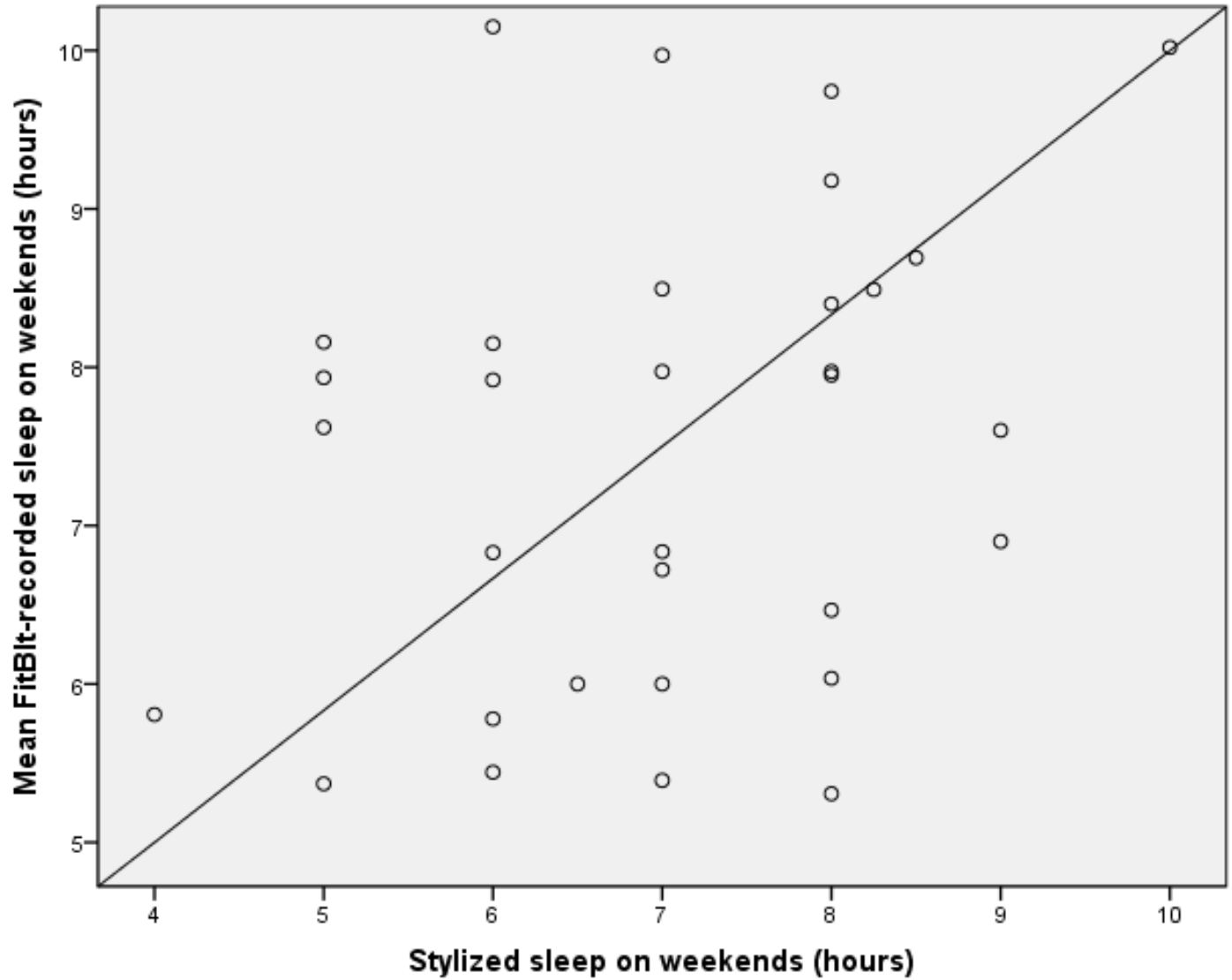
Good agreement

ICC = .62
(CI .35-.79)
 $\rho = 0.59$

$p < .05$

Stylized and FitBit-Recorded Sleep (weekend)

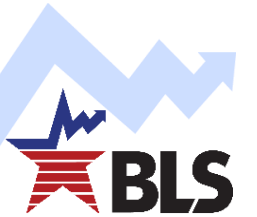
Stylized Question: Thinking about the past weekend [fill dates] how many hours of sleep did you get on average each night?



Poor agreement

ICC = .30
(CI -.05 -.58)
rho = 0.28

N.s.

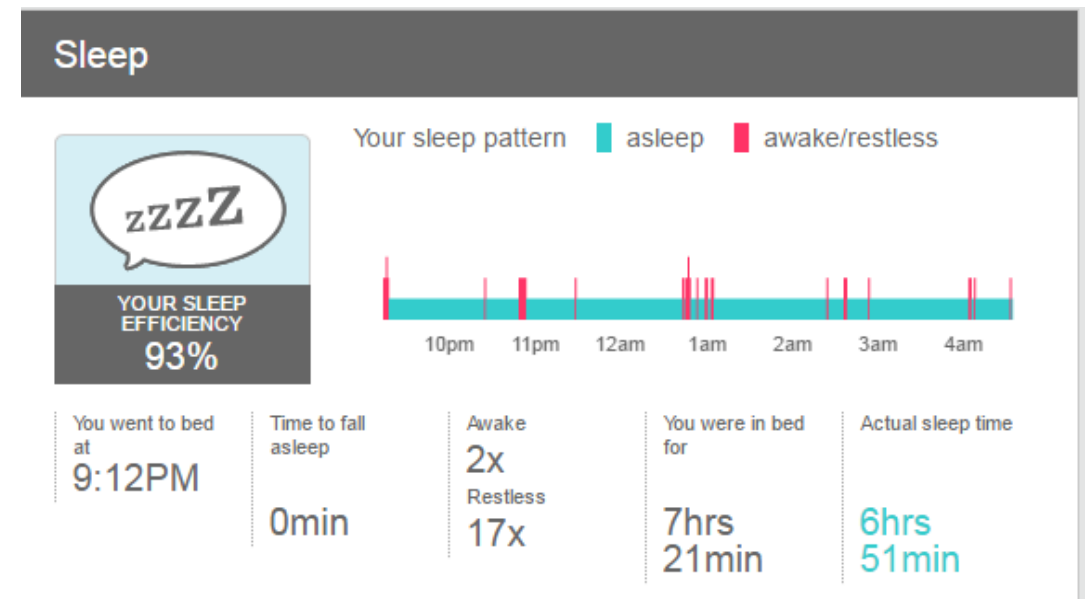


Summary: Agreement Across Sleep Duration Measures



Debriefing

- 18/35 thought the Fitbit sleep data was about right
 - ▶ 10 thought it underestimated
 - ▶ 7 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 FitBit each night (n=8)
- ▶ Only wore FitBit one weeknight of the week (n=7)
- ▶ Lost FitBit (n=2), but one found & returned it later!
- ▶ FitBit fell off during the night (n=1)

Conclusions



- Measurement error in self-reports of sleep
- Objective sleep measures might fall in between diary and stylized reports of sleep
 - ▶ Agreed more with diary vs. stylized measures
 - ▶ FitBit had its own set of measurement and user error
- FitBit-recorded data was a useful memory aid
- Qualitative and quantitative pretesting techniques can build off one another to provide a more complete picture of survey measurement error sources

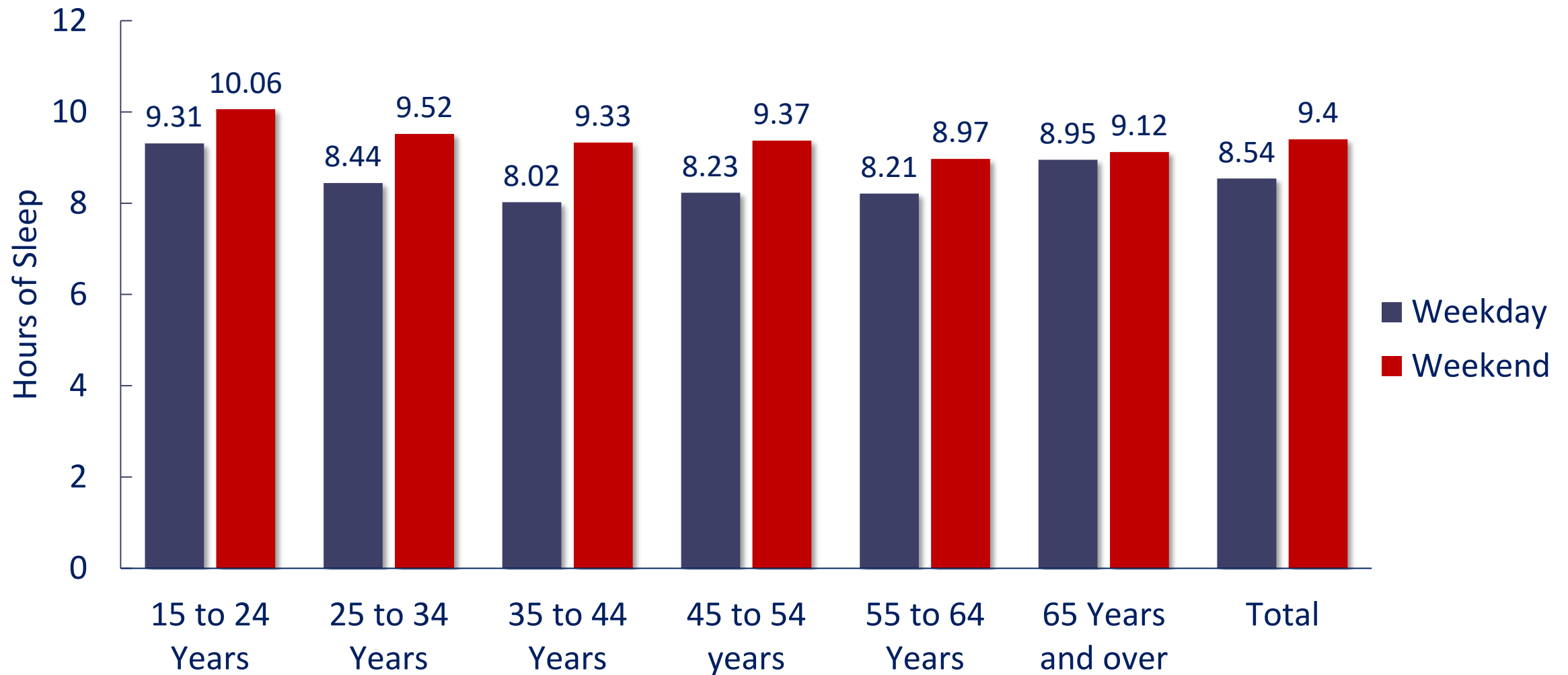
Contact Information

Robin Kaplan, Brandon Kopp, and Polly Phipps
Bureau of Labor Statistics
Office of Survey Methods Research (OSMR)
FedCASIC 2019

E-mail:
Kaplan.Robin@bls.gov

Phone:
(202) 691-7383

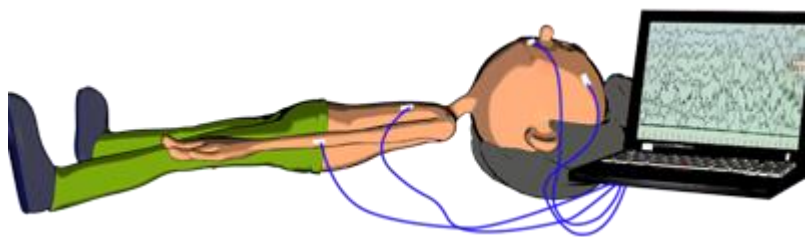
Sleep Duration in the ATUS



* Data are annual averages for 2014.



Objective Sleep Measures



Polysmnography (PSG)

€€€



Sleep Actigraphy

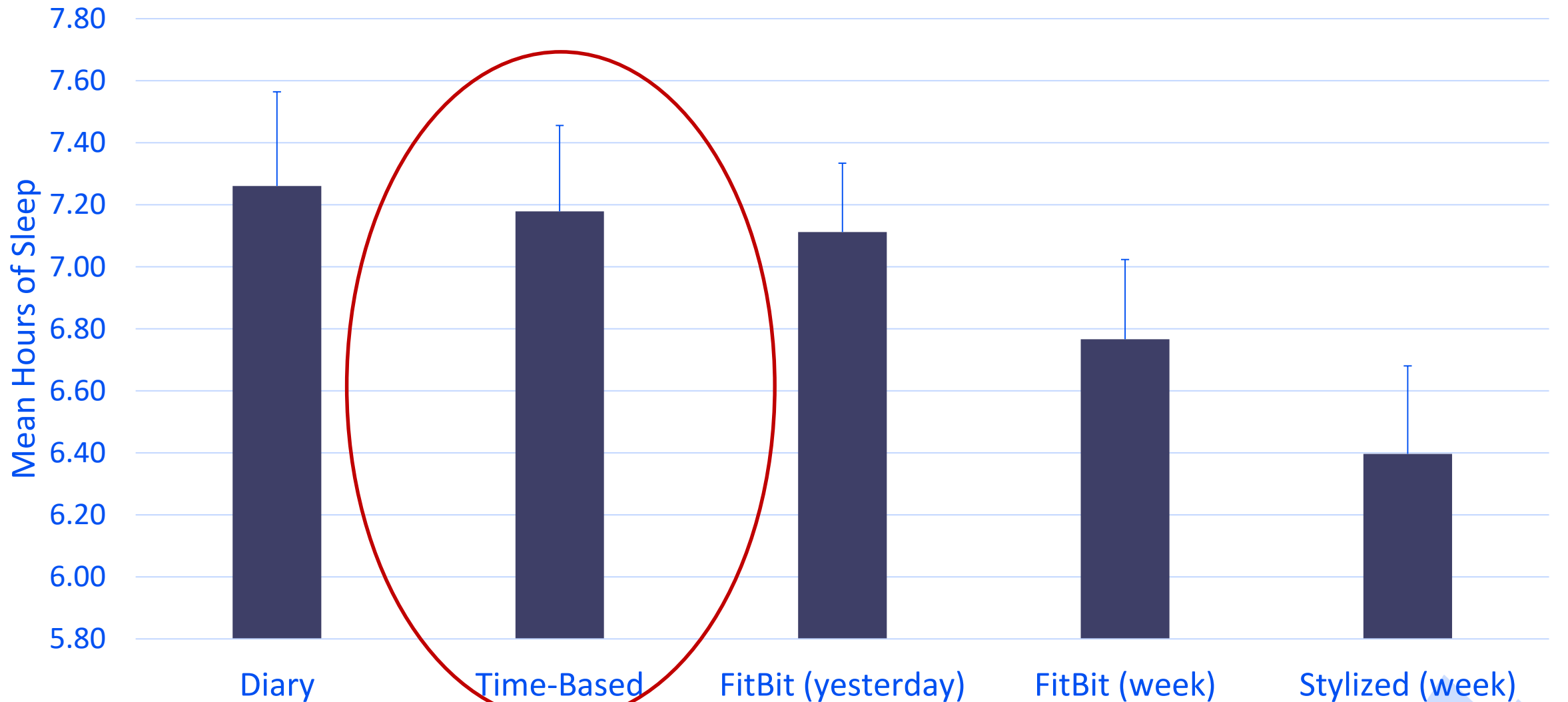
€€



FitBit

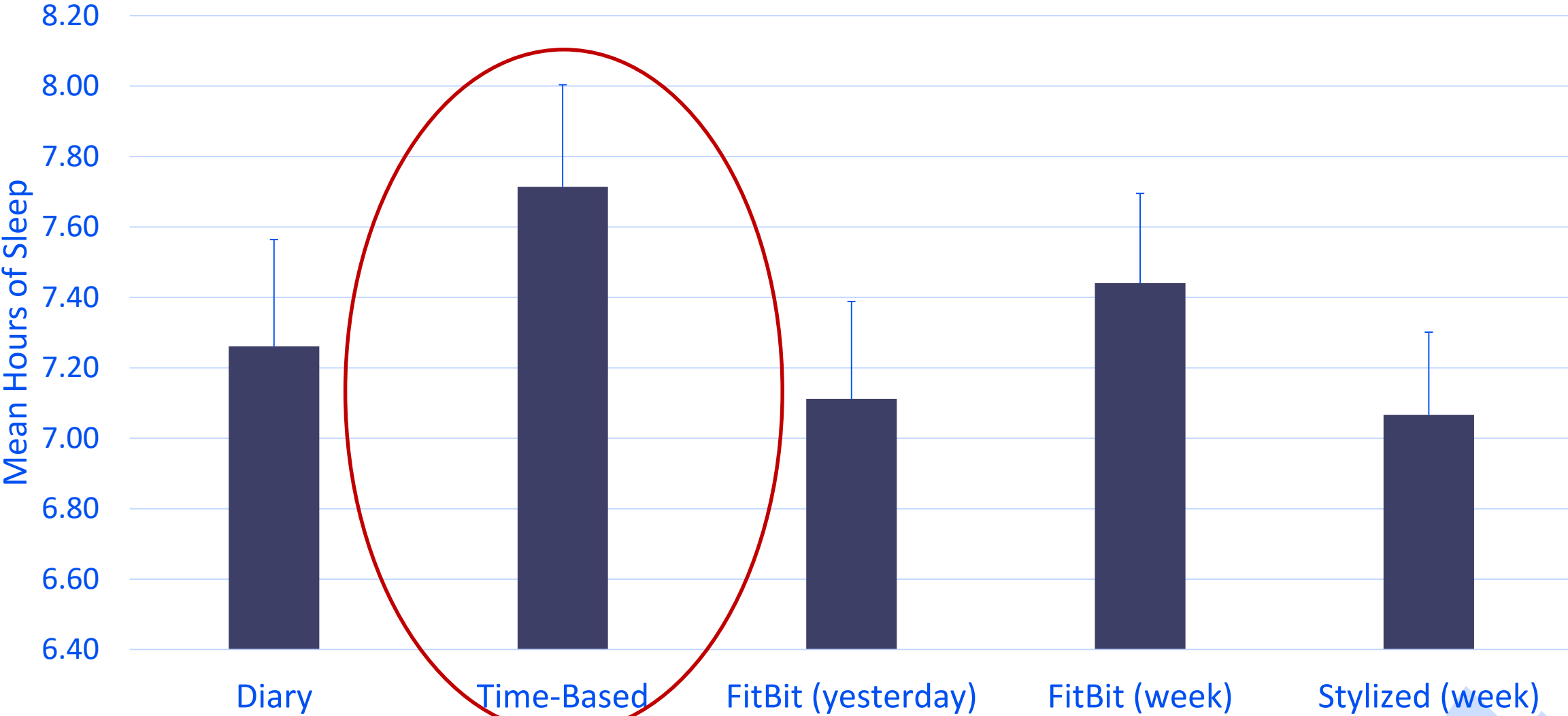
€

Results (Weekdays)



Diary, Time-Based, and FitBit (yesterday) exceeded Stylized (all ps < .005)

Results (Weekends)

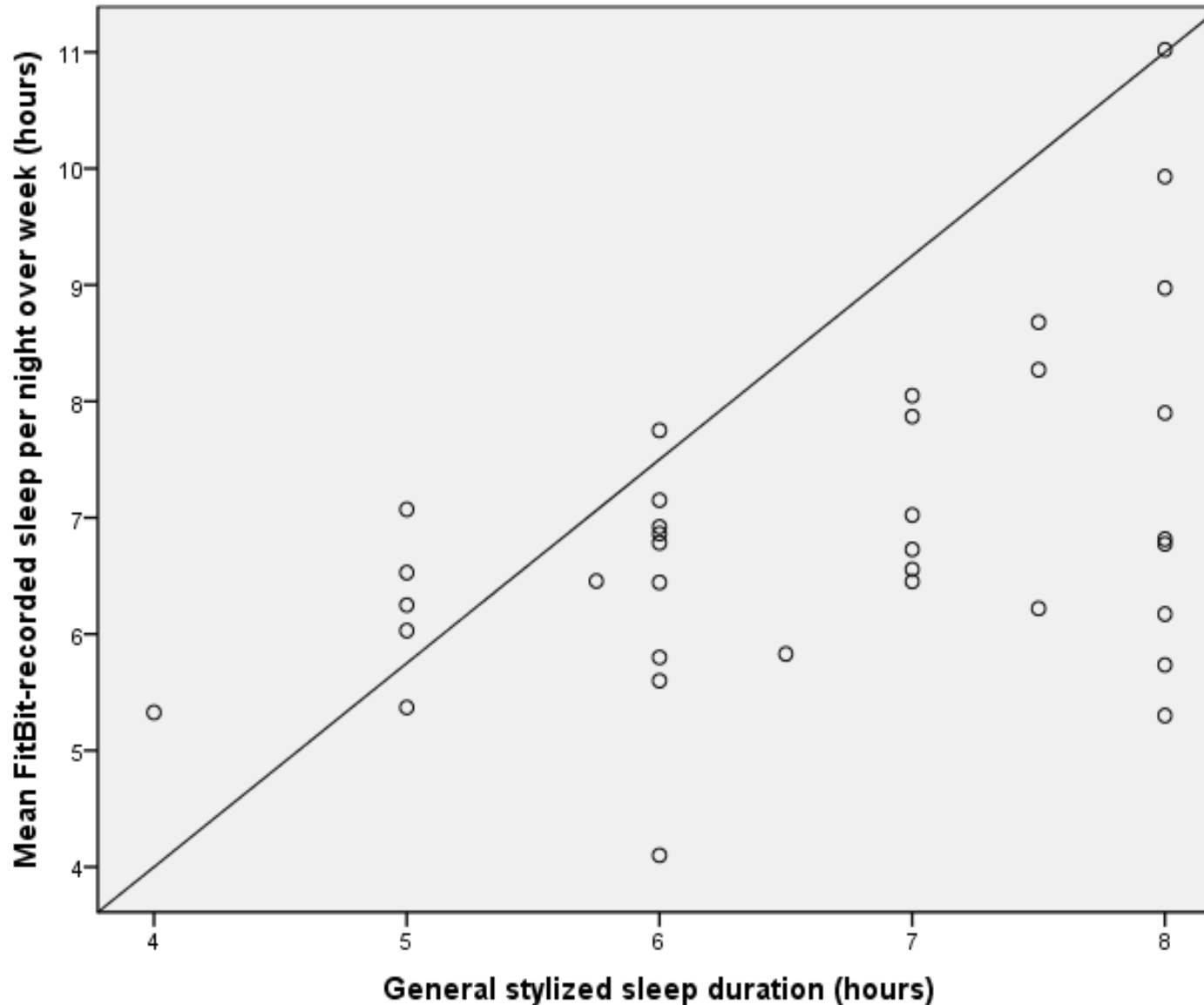


No significant differences



General Stylized and FitBit-Recorded Sleep (weekdays + weekends)

General Stylized Question:
On average, how many hours of sleep do you get per night?

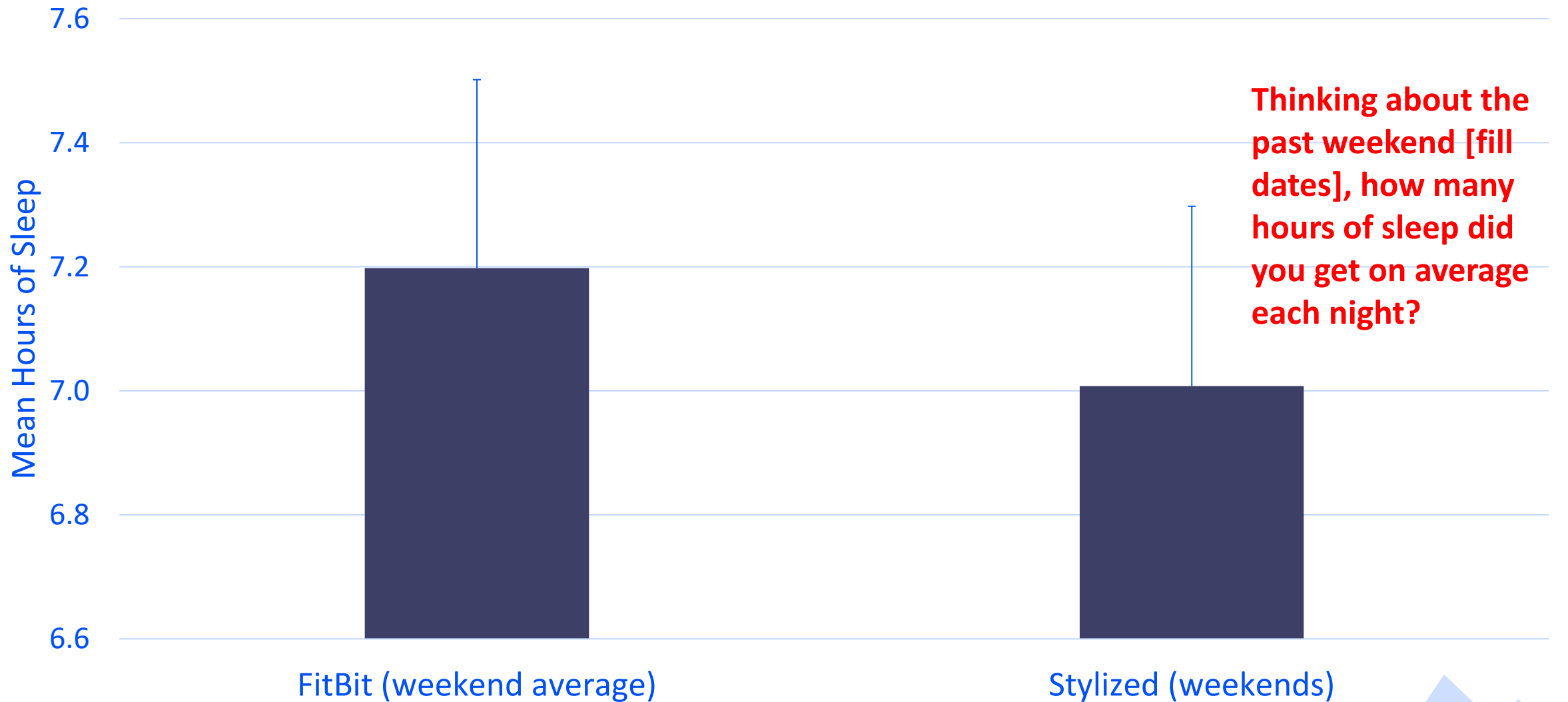


$\rho = 0.37$
ICC = .45
(CI .15-.68)

$p < .05$

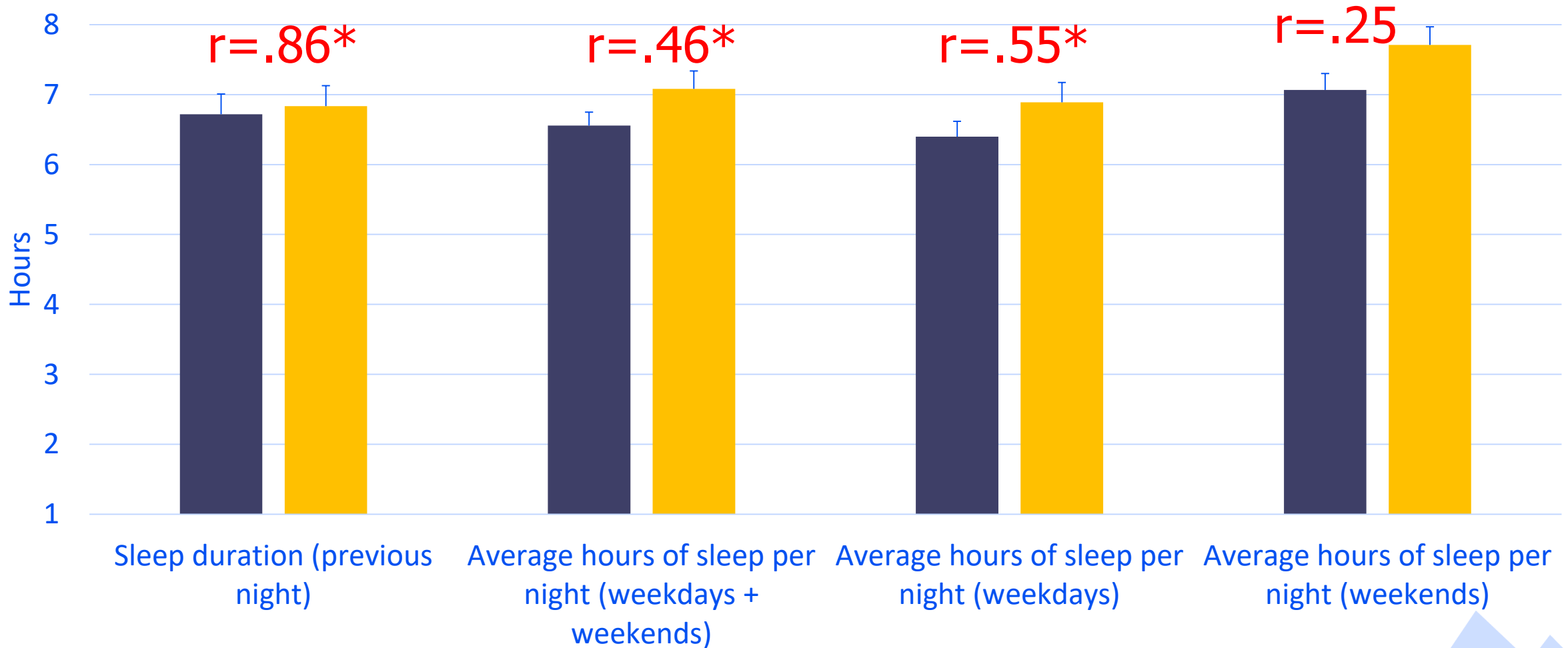
Fair agreement

Sleep Duration Measures (Weekends)



No significant difference

Self-Reported Sleep versus FitBit-Recorded Hours in Bed



ICC cutoffs

- Less than 0.40—Poor.
- Between 0.40 and 0.59—Fair.
- Between 0.60 and 0.74—Good.
- Between 0.75 and 1.00—Excellent.



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