Article

Expert Demand for Consumer Sleep Technology Features and Wearable Devices

Jaime K Devine[1*], Lindsay P. Schwartz[1], Jake Choynowski[1], Steven R Hursh, PhD[1,2,]

[1]   Institutes for Behavior Resources, Baltimore, MD, USA 21218; jdevine@ibrinc.org; lpschwartz@ibrinc.org (L.P.S.); jchoynowski@ibrinc.org (J.C.); shursh@ibrinc.org (S.R.H.)

[2]   Johns Hopkins University School of Medicine, Baltimore, MD, USA 21205; shursh@ibrinc.org (S.R.H.)

*   Correspondence: jdevine@ibrinc.org; Tel.: +1-410-752-6080 (ext. 132)

Abstract: Global demand for sleep-tracking wearables, or consumer sleep technologies (CSTs), is steadily increasing. CST marketing campaigns often feature a scientific component, but the scientific relevancy and monetary value of CST features within the sleep research community remains unquantified. Sleep medicine experts were recruited through social media and nonprobability sampling techniques to complete a survey identifying sleep metrics and device features that are most desirable to the scientific community. A hypothetical purchase task (HPT) estimated economic valuation for devices with different features by price. Forty-six (N=46) respondents with an average of 10±6 years' experience conducting research in real-world settings completed the online survey. Total sleep time was ranked as the most important measure of sleep followed by objective sleep quality while sleep architecture/depth and diagnostic information were ranked as least important. Experts preferred wrist-worn devices that could reliably determine sleep episodes as short as 20 minutes. Economic value was greater for hypothetical devices with longer battery life. These data set a precedent to determine how scientific relevance of a product impacts the potential market value of a CST device. This is the first known attempt to establish consensus opinion or economic valuation for scientifically-desirable CST features and metrics using expert elicitation.

Keywords: behavioral economics; wearables; consumer sleep technology; Internet of Things; economical survey; expert elicitation

1. Introduction

Wearables are a hallmark of the Internet of Things (IoT), but sleep researchers have been using a precursor to wrist-worn wearables called actigraphy since before the "birth" of the internet in 1983 [1-3]. Actigraphy has been used to determine sleep-wake behavior using predominantly activity data since the 1970s whereas internet-enabled applications and wearables began to hit the market in the mid-2010s [1,2,4]. A 2015 Journal of Clinical Sleep Medicine article coined the term used to describe publicly-available computer-based systems that aim to monitor or improve individual sleep behavior defined as "consumer sleep technologies (CSTs)" [5]. This term is widely used by the sleep research community to describe wearables or non-wearable sleep tracking technology [1]. The authors noted in 2015 that there was "scant literature discussing these technologies", providing only two examples of previous studies [4-6]. Since its publication in 2015, however, "*Consumer Sleep Technologies: A Review of the Landscape*" has been cited 167 times according to Google Scholar. The literature is no longer scant.

Enough reviews, statements, editorials, validation studies, and research findings focused on CSTs have been published that it is impossible to cite them all. An underlying message across these articles has been that CSTs are becoming a part of the research landscape, but that technologies cannot be taken seriously until they have been shown to be accurate compared to traditional sleep measurements [7-14]. A 2021 study by Chinoy et al. compared the performance of seven recently-developed CST devices

against polysomnography (PSG) and found that CSTs exhibited high performance in detecting sleep that rivaled, and in some cases surpassed, the accuracy of research-grade actigraphy [15]. Relatedly, an in-depth review of the history and future of multisensory sleep-tracking wearable technology was recently published by Lujan, Perez-Pozuelo, and Grandner in the journal *Frontiers in Digital Health* [1]. These publications, as well as countless other scientific analyses comparing CSTs against traditional scientific sleep measurement systems, indicate a turning point in the viability of CSTs for research purposes.

The scientific community cannot ignore the benefit of potentially superior fieldable sleep measurement systems. However, as the name suggests, CSTs have been designed for the everyday consumer rather than as a reliable scientific tool. The design and features of CST wearables appear to be driven by what will influence the buying decision of the average consumer. This approach assumes that what the consumer values is grounded in good empirical evidence as to what information about sleep is truly important to monitor. To address that assumption, we set out to ask academic and industry sleep researchers, who are well-educated about the value of sleep monitoring, what features of a wearable are most important in a CST wearable. The study described here was designed to answer the following specific questions:

1. Which metrics of sleep quantity and quality do experts in the field believe are most important for a CST wearable to measure?
2. What wearable design features are most important for successful tracking of sleep in the real-world from the perspective of experts who conduct such studies?
3. How much economic value do experts place on a CST wearable that has the most desirable sleep metrics and design features?

Establishing the economic value of CST wearable features in the scientific context is an on-going, multi-step project being conducted as a collaboration between the Institutes for Behavior Resources (IBR) Operational Fatigue and Performance group and the Behavioral Economics group. Behavioral economics is a field of science that applies behavioral science within economic frameworks to explain decision-making behavior [16-18]. In order to assess the economic value of CST features that are desirable to sleep researchers, however, we first needed to establish which features are, in fact, desirable to sleep researchers in economic terms.

The first step in the project was to conduct an online survey to see what wearable features scientists with a professional interest in measuring real-world sleep consider most important [19]. Survey items were geared toward identifying discrete design features as well as a hypothetical purchase task (HPT), a validated behavioral economic demand procedure for evaluating demand across a range of hypothetical circumstances. This procedure can provide information about sensitivity to price for a particular device or feature (e.g., how purchasing behavior may change as a function of cost) and provide an economic valuation of those features in terms of demand elasticity. The target recruitment population for this survey were scientists and industry professionals who routinely conduct research related to human sleep physiology or behavior in real-world environments, i.e., sleep research experts.

One concern with an expert survey is that the sample population may not be adequately representative of the expert community [20]. Adequate sampling requires the determination of a sufficient sample size based on an estimation of the size of the overall population [21]. Experts can be recruited using nonprobability sampling techniques, such as convenience sampling [22,23] or purposeful sampling [24-26]. "Snowball" sampling, in which participants identify other potential participants, is another nonprobable sampling technique which can help increase the recruitment of experts for survey participation [20,27,28]. This paper describes the efforts taken to ensure that

respondents were representative of the global sleep expert population and presents a rank ordering of researcher preferences for CST device features and metrics.

The goal of this manuscript is to establish experts' consensus opinion of the economic value of desirable CST wearable features for use in real-world sleep research environments. To our knowledge, this is the first report of an examination of preferred CST wearables features solicited from sleep experts and a behavioral economic analysis of the value this group assigns to those features. These findings will next be used to assess the extent to which currently available devices meet these criteria and to inform the development of a behavioral economic survey in a general consumer population for CST wearables to determine the added value conferred by scientific validation and endorsement.

2. Materials and Methods

Professional opinions from sleep medicine experts were elicited to identify what metrics and device features for measuring sleep outside the laboratory are most desirable to the scientific community. Potential respondents were recruited actively through direct contact and passively through social media on Twitter (www.twitter.com) and LinkedIn (www.linkedin.com) social media platforms and through scientific presentations which described the scope of the project to the target audience. Potential respondents were actively recruited through email based on the results of a literature review conducted using the biomedical literature search engines Pubmed.gov (https://pubmed.ncbi.nlm.nih.gov/) and scholarly literature web search engine Google Scholar (https://scholar.google.com/).

Search criteria for the literature review included that the article had been peer-reviewed, published before July 2021, and included a combination of the terms 'sleep,' AND 'environment' OR 'operational' OR 'real-world' OR 'ecological' AND/OR 'wearable' OR 'consumer sleep technology' OR 'device' in the topic or title fields. Each returned article was manually scanned for relevancy. Articles which included a description of methods of active data collection in a real-world environment or simulated real-world environment were considered relevant. Review articles, meta-analysis, papers focused on device engineering/development, or opinion pieces such as editorials were not sufficient for inclusion. Authors who had published at least two or more relevant papers were subsequently considered subject matter experts and were contacted via blind carbon copy (bcc) email with a brief written explanation of the purpose of the research and a link to the online survey. Recruitment terminology on all platforms (social media, scientific presentation, and email) included a request to share the survey with any interested colleagues in relevant fields in order to create a snowball effect and reach a broader range of potential respondents.

The survey was hosted through the online tool Qualtrics (www.qualtrics.com) between April and July 2021. The voluntary anonymous survey was composed of 42 questions grouped in 5 sections. The first section contained 5 questions that focused on identifying each respondent's background and research experience. The second section contained 10 questions about respondents' typical sample population and study design. The third section contained 13 questions asking about device preferences, and the fourth section asked respondents to rate their agreement with 6 statements. The fifth section contained 9 questions regarding economic demand for devices with varying features and price points. Respondents were able to provide comments in a text box up to a 20k character limit at the end of the survey. This study was approved by the Salus Institutional Review Board and these analyses were conducted in accordance with the Declaration of Helsinki.

An annotated list of survey items and hypothetical devices for discussion in this paper is outlined in Table 1. Respondents could select the best fitting option from a multiple-choice list for Q1-Q12 or provide a write-in response if no option described them. Respondents were asked to rank by order of importance (high importance,

medium importance, or low importance) a list of features or metrics related to the question topic for Q13-15. Respondents could additionally provide and rank a write-in response.

All data were exported from Qualtrics as an Excel file and subsequently analyzed using Excel 2013. In-depth statistical testing was not appropriate for these analyses due to the nature of the survey and the small sample size. Responses for Q13-15 were weighted by level of importance. The Excel Rank function was used to calculate weighted mean rank order for Q13-Q15 items. The hypothetical CST devices used for the HPT are briefly described in Table 2.

Table 1. Survey Questions and Number of Responses.

Question	Number of Responses
Research Background and Experience	
Q1. Do you conduct human subjects research related to sleep in real-world environments/outside a controlled laboratory environment? (*Only respondents who selected Yes were able to complete the rest of the questions*)	55
Yes	46
Q2. How many years' experience do you have conducting human sleep research in real-world environments?	44
Q3. Which organization best describes your research affiliation?	43
Q4. In what region are you/your research based?	42
Q5. Which category best describes the population whose sleep you study?	40
Q6. Which category best describes the focus of your research?	36
Device Preferences: Multiple Choice	
Q7. Where is your preferred placement for a fieldable device or instrument to collect sleep data?	32
Q8. What do you consider to be the most important time scale for measuring sleep for your research in general?	
Q9. What is the most appropriate method for determining actual sleep onset/offset in real-world environments?	33
Q10. Do you collect data related to napping or fragmented sleep? (yes/no)	29
Yes	24
Q11.What is the most appropriate minimum period of inactivity that could reliably be considered a nap? (only respondents who selected Yes on Q10 received this question)	23
Q12. What is your preferred continuous observation period or window for collecting data on real-world sleep?	32
Device Preferences: Rank Order of Importance	
Q13. Which information about sleep do you consider most important to your research?	33
Q14. Which features of a fieldable device or instrument are/would be most important to facilitating data collection for your research?	30
Q15. Which factors related to devices or instruments are most important to limiting your observation period/data collection window?	33

Table 2. Device Descriptions for the Hypothetical Purchase Task.

	Device A	Device B	Device C

	• Accelerometer • Light exposure • 30 second epochs • 30 days continuous battery life	• Accelerometer • Light exposure • Heart rate detection • Oxygen saturation • 4 days continuous battery life	• Accelerometer • Light exposure • Heart rate detection • Oxygen saturation • 30 days continuous battery life

Specifications	• 30 days on-device data storage
	• Data is downloaded from device via USB
		• 100 days cloud data storage
		• Remote data download/Bluetooth
		
		Sleep depth estimation (Light, Deep, REM) which has been validated against PSG
			• Extraction mode settings:
			  • 30 days on-device storage via USB download
			  • 100 days cloud data storage via Bluetooth
			• Sleep depth estimation (Light, Deep, REM) which has been validated against PSG

Data Features	• 30 days on-device data storage
	• Data is downloaded from device via USB
		• 100 days cloud data storage
		• Remote data download/Bluetooth
		
		Sleep depth estimation (Light, Deep, REM) which has been validated against PSG
			• Extraction mode settings:
			  • 30 days on-device storage via USB download
			  • 100 days cloud data storage via Bluetooth
			• Sleep depth estimation (Light, Deep, REM) which has been validated against PSG

Features Consistent Across All Devices	• 30 second epochs
	• Rechargeable battery
	• Water-resistant to 50 m
	• Dual function as a watch and alarm clock
	• Event marker
	• Optional feedback mode
	• Companion software which allows researchers to score raw activity data
	• CSV export of epoch-by-epoch (EBE) data
	• CSV export of summary data averaged across all study days:
	  • Bedtime in 24-hour clock time
	  • Waketime in 24-hour clock time
	  • Total Sleep Time (TST) in minutes
	  Sleep Efficiency (SE) as a percentage		

Participants' data were first analyzed with algorithms used to determine non-systematic demand curve data [29]. Four demand curves were excluded as data were only input for a single price (1 curve for Device A and B and 2 curves for Device C). Demand curve data were analyzed using a recent extension of the exponential demand model [30] specifically designed to handle zero demand data using guidelines and equations from Gilroy et al., (2021) [31]. With this equation, Q_0 is an estimate of the maximum level of demand (the number of devices one would purchase), and α is an estimate of the rate of change in elasticity normalized to the transformed maximum level of demand, Q_0. A demand curve template for GraphPad Prism 8.0 available from the Institutes for Behavior Resources (www.ibrinc.org) was used to fit the pooled consumption data and

estimate the two parameters. Post Hoc Extra Sum of Squares F-tests were used on the pooled consumption data to determine whether the demand elasticity rate parameter (α) differed between devices. We also report Essential Value (EV) which is proportional to the inverse of α (EV = $1/(100 \times \alpha)$) and allows for a simple way to understand the rate of change in elasticity – lower rate of change in elasticity is denoted by smaller α values and indicates higher EV or higher resistance to the effect of price [30]. Taking the inverse of α ensures that a higher value is represented by a higher number. P*max* was also calculated for each curve; P*max* denotes the price at which demand becomes elastic and monetary expenditure would be maximal and was calculated using an Excel solver tool that uses α and Q*0* to estimate P*max* (https://ibrinc.org/behavioral-economics-tools/).

3. Results

3.1. Respondents

3.1.1. Recruitment Exposure

The total estimate of exposure of potential respondents through all platforms was N=14,057. The vast majority of exposure occurred through the social media website Twitter (N=11,117). Between April to July 2021, a total of 37 Twitter users clicked the research survey link which had been embedded in recruitment tweets and 14 Twitter users retweeted the recruitment tweets. Estimated exposure through scientific presentations was approximately 142 attendees. Additionally, 101 potential respondents were directly recruited through email.
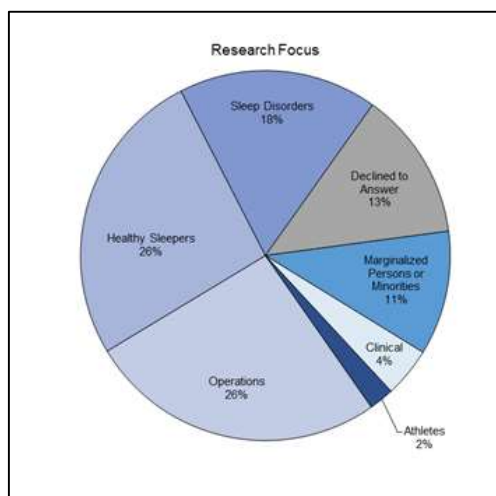
3.1.2. Demographics

Seventy-six (76) individuals navigated to the Qualtrics survey website. Of these, 55 respondents indicated whether they conducted research related to sleep in real-world environments (see Table 1 above). Nine (N=9) respondents indicated that they did not conduct research in this field and were excluded from further participation, leaving a total sample population of N=46. Respondents did not need to complete the entire survey in order to be included in subsequent analysis. All respondents completed between 5-100% of the survey. Sixty-three percent (63%; N=29) of respondents completed 100% of the rank order survey questions and 39% (N=18) of respondents completed the HPT.
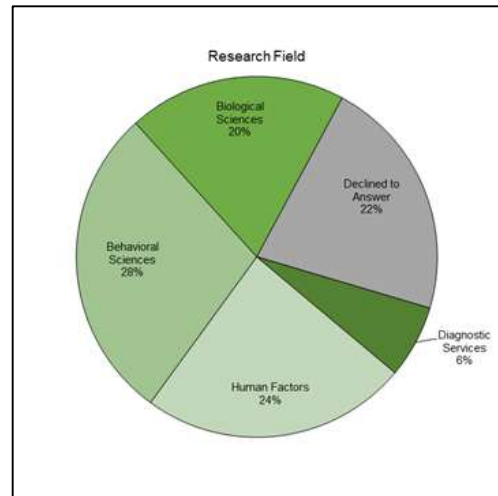
Figure 1A depicts respondents' years of experience by region: Africa; Asia; Europe; North America; Oceania; South America; and organization type: Academic or Education Institution; For-Profit Organization or Business; Government; Hospital or Clinical Laboratory; Non-Profit Organization. The majority of respondents (N=26) were geographically located in North America and conducted research in an academic or education institution (N=30). Respondents had an average of 11 years' experience conducting research (Range: 1-27 years; Mode: 10 years). Figure 1B shows the distribution of respondents by research focus across the following categories: 1) healthy sleepers with conventional sleep patterns; 2) sleep disorders; 3) operational environments, such as transportation, shiftwork, or the military; 4) marginalized or underrepresented groups, including racial, ethnic, and gender minorities; 5) other. One (N=1) respondent indicated that their research focus pertained to ultra-endurance sports athletes, which constitutes its own category in Figure 1B. Six (N=6) respondents did not provide an answer. Figure 1C shows the distribution of respondents by research field: 1) biological sciences; 2) behavioral or social sciences; 3) human factors or ergonomics; 4) diagnostics services. No respondent indicated that they worked in a field other than these categories; 10 respondents did not provide an answer.
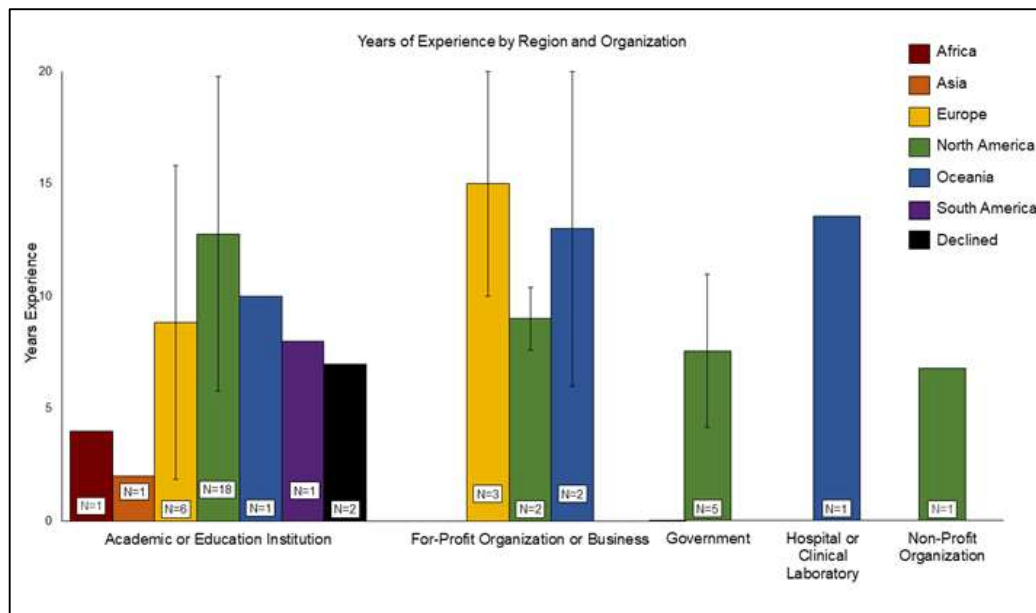
(**a**)

(**b**)   (**c**)

Figure 1. Respondent Demographics by Years' Experience, Region, and Research: (**a**) Respondents' years of experience (y-axis; mean ± standard deviation) plotted by organization (x-axis) and global region (legend). The number of respondents (N) per category is indicated by the white inset box. Percentage of respondents by (**b**) research focus and (**c**) research field are depicted in pie charts.

3.2. Device Preferences

Responses to Q7 -Q12 are depicted in Figure 2. Percentages are shown by total number of respondents per question rather than total number of survey respondents. In brief, a 66% majority of respondents (21/32) preferred devices which are worn on the wrist for sleep measurement (Q8) and a 49% majority of respondents (16/33) considered epoch-by-epoch/minute-by-minute to be the most important time scale for measuring sleep (Q8). On Q9, 52% (15/29) believed that the most appropriate method for determining sleep onset/offset was through a combination of brain activity, motor activity (e.g., activity counts), and peripheral biometric measures (e.g., heart rate, oxygen saturation). Eighty-three percent (83%; 24/29) of respondents indicated that they collected data

related to napping or sleep fragmentation (Q10). Only these 24 respondents were per-mitted to answer Q11. On Q11, 52% (12/23) indicated that the most appropriate mini-mum period of inactivity that could be considered a nap was less than or equal to 20 minutes. Another sizeable fraction (44%) of respondents selected a time period between 20 to 40 minutes. No respondents selected a time period greater than 60 minutes for Q11. For Q12, a 44% majority of respondents (14/32) preferred a continuous observation win-dow between 4 to 14 days long.
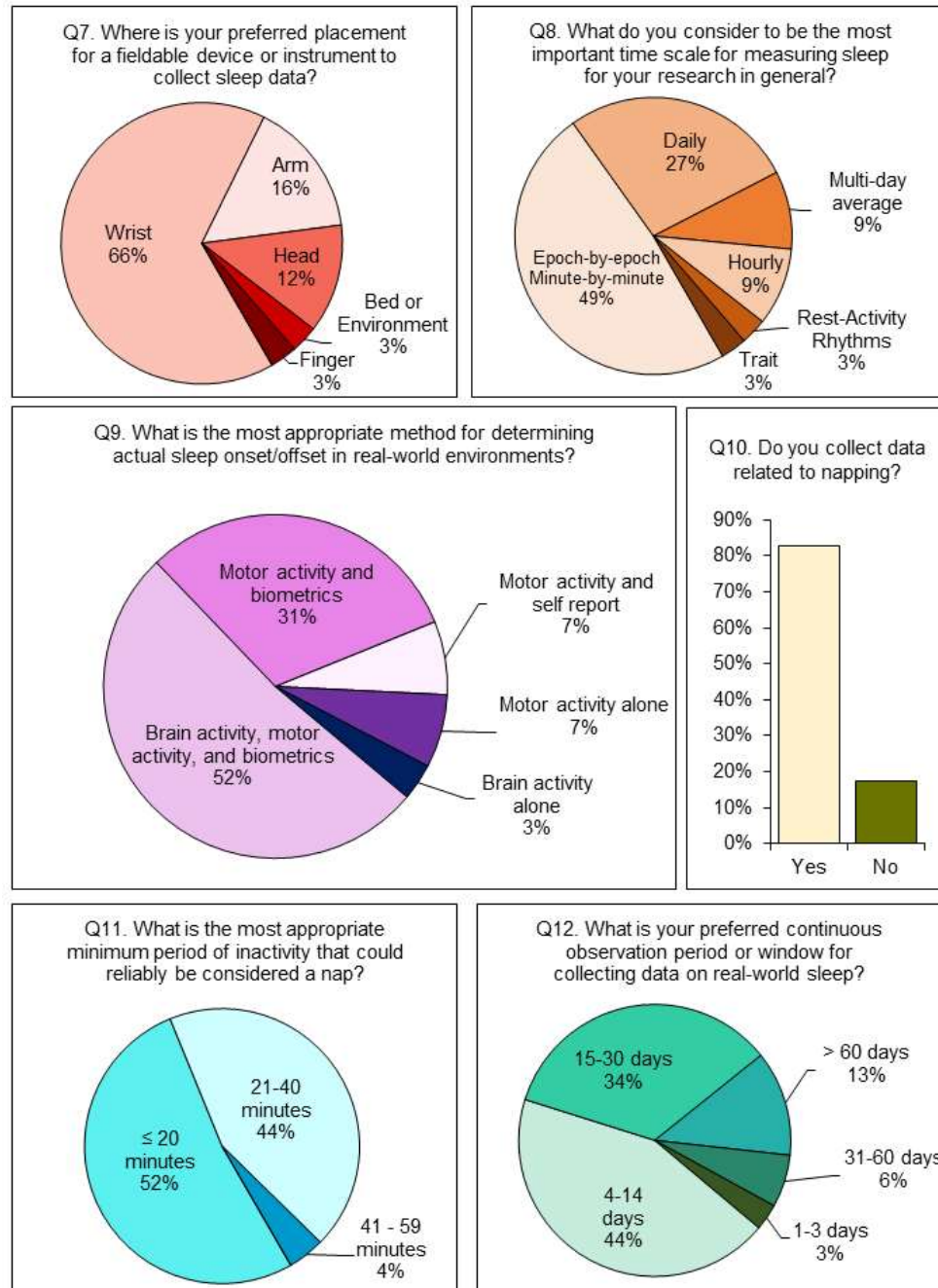

Figure 2. Device Preferences: Multiple Choice: Responses to Q7-Q12 regarding device preferences by number of respondents per question. Pie slices indicate multiple-choice options and percent of respondents per question who selected this response. Multiple choice options which were not selected by any respondent (0%) are not depicted.

Rank order responses are depicted in Figure 3. For Q13, Total Sleep Time (TST) received the highest ranking for information about sleep that respondents considered important to their research followed by objective sleep quality, time in bed (TIB), and subjective sleep quality. Measures of sleep architecture or sleep depth and diagnostic information (e.g., apnea-hypoxia index; AHI) were ranked the lowest out of provided categories. One respondent provided a write-in response, indicating that social activity timing was of high importance to their research. Regarding device features which facilitated data collection (Q14), the ability to differentiate between actual sleep and "false sleep" (i.e., a period of low activity that mimics sleep onset but is not a sleep episode) was most frequently ranked as having high importance, followed by data security. The ability for subjects to self-report information other than sleep and the capacity to provide feedback or an intervention were ranked lowest. For Q15, the top ranked factor which respondents felt limited their observation period or window of data collection was battery life. Logistics, such as receiving approval for study procedures, was ranked lowest. Four (4) respondents provided additional comments about the use of CSTs for research. End-of-survey comments are included in Appendix A.
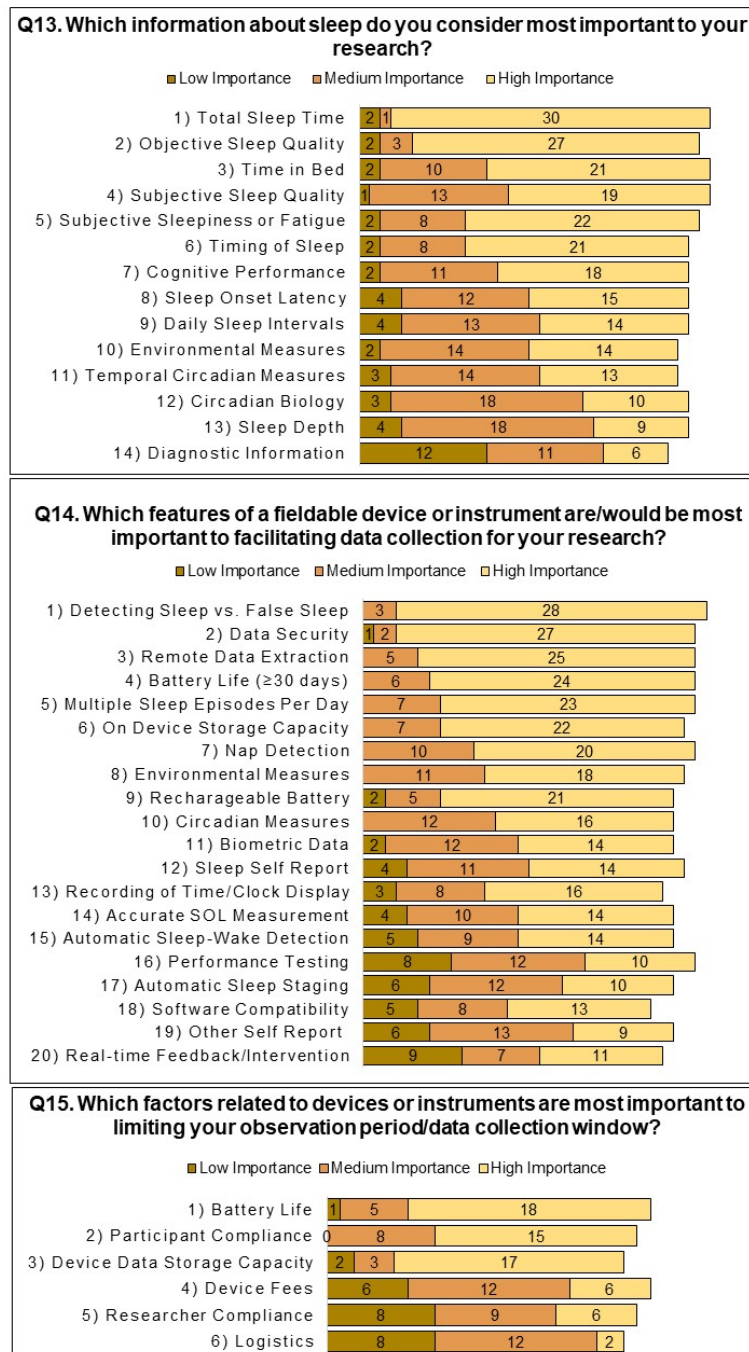
Q13. Which information about sleep do you consider most important to your research?

■ Low Importance ■ Medium Importance □ High Importance

Item	Low	Medium	High
1) Total Sleep Time	2	1	30
2) Objective Sleep Quality	2	3	27
3) Time in Bed	2	10	21
4) Subjective Sleep Quality	1	13	19
5) Subjective Sleepiness or Fatigue	2	8	22
6) Timing of Sleep	2	8	21
7) Cognitive Performance	2	11	18
8) Sleep Onset Latency	4	12	15
9) Daily Sleep Intervals	4	13	14
10) Environmental Measures	2	14	14
11) Temporal Circadian Measures	3	14	13
12) Circadian Biology	3	18	10
13) Sleep Depth	4	18	9
14) Diagnostic Information	12	11	6

Q14. Which features of a fieldable device or instrument are/would be most important to facilitating data collection for your research?

■ Low Importance ■ Medium Importance □ High Importance

Item	Low	Medium	High
1) Detecting Sleep vs. False Sleep	3		28
2) Data Security	1	2	27
3) Remote Data Extraction	5		25
4) Battery Life (≥30 days)	6		24
5) Multiple Sleep Episodes Per Day	7		23
6) On Device Storage Capacity	7		22
7) Nap Detection	10		20
8) Environmental Measures	11		18
9) Rechargeable Battery	2	5	21
10) Circadian Measures	12		16
11) Biometric Data	2	12	14
12) Sleep Self Report	4	11	14
13) Recording of Time/Clock Display	3	8	16
14) Accurate SOL Measurement	4	10	14
15) Automatic Sleep-Wake Detection	5	9	14
16) Performance Testing	8	12	10
17) Automatic Sleep Staging	6	12	10
18) Software Compatibility	5	8	13
19) Other Self Report	6	13	9
20) Real-time Feedback/Intervention	9	7	11

Q15. Which factors related to devices or instruments are most important to limiting your observation period/data collection window?

■ Low Importance ■ Medium Importance □ High Importance

Item	Low	Medium	High
1) Battery Life	1	5	18
2) Participant Compliance	0	8	15
3) Device Data Storage Capacity	2	3	17
4) Device Fees	6	12	6
5) Researcher Compliance	8	9	6
6) Logistics	8	12	2

Figure 3. Device Preferences: Rank Order: Mean rank order of responses to Q13-Q15 regarding device preferences by level of importance. Items are listed on the y-axis by weighted rank, with number 1 corresponding to higher importance ranking. Bars depict the number of responses by level of importance (low importance in dark gray, medium importance in gray, high importance in light gray) for each item.

3.3. Behavioral Economic Demand

Demand curves were created for the hypothetical devices described in Table 2. Demand curves are depicted in Figure 4. The Gilroy et al. (2021) model of demand [31] fit the data well, with an average R^2 of 0.9847. Post Hoc Extra Sum of Squares F-tests determined a significant difference between curves for the three devices ($F_{(2, 447)}=6.01$, $p=0.003$). Further examination showed that Device B had significantly lower Q_0 than

Device A and C, indicating that when the devices are free, participants would 'purchase' fewer units of Device B than Device A and C. Additionally, F-tests on α values indicated demand for Device B was significantly more elastic than demand for Devices A and C. The essential values for Device A, B, and C were 705, 367, and 935, respectively. In rank order, Device C maintained the highest purchasing level across prices while Device B maintained the lowest.
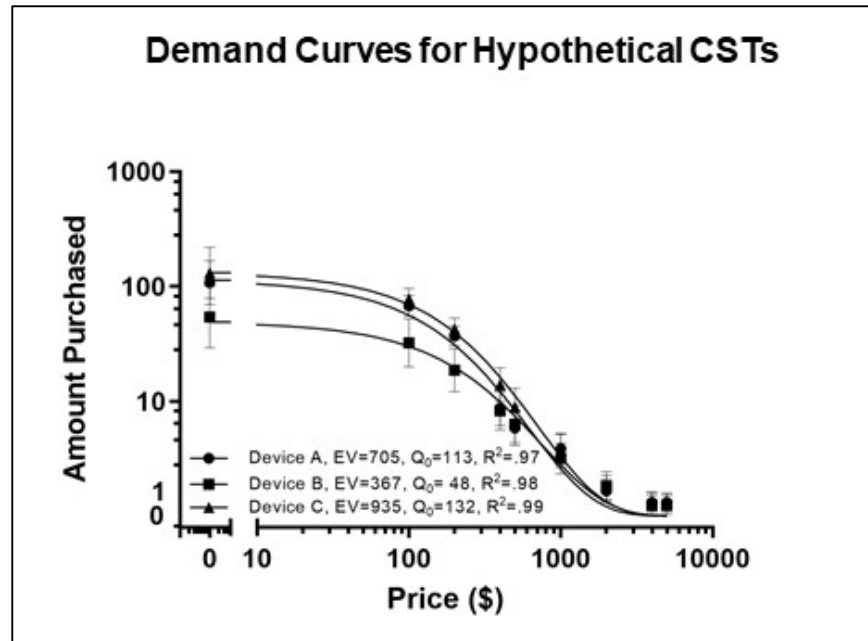

Figure 4. Demand Curve for Hypothetical CST Devices: Group mean demand curves for Device A, Device B, and Device C. Essential Value (EV) indicates the consumption level to be expected at a particular price; Q0 is an estimate of the maximum level of demand; R2 represents the coefficient of determination. Consumption data is shown in Inverse hyperbolic sine (IHS) units.

4. Discussion

The main aim of the survey was to establish a reliable consensus opinion from sleep research professionals about preferred device features for measuring sleep outside the laboratory in the context of economic demand. This is important not only to provide guidance to CST or wearables manufacturers, but also to facilitate innovation within sleep research methodology. The reliability of the survey results hinges on the assumption that a sufficient number of subject matter experts provided responses. Survey completion was done anonymously to ensure that respondents would feel comfortable providing honest responses, but may create skepticism about the expertise of unknown respondents. While no identifying information was collected, respondents had to indicate that they conducted sleep research related to human sleep in real-world environments in order to complete the survey. Expertise may be considered a matter of opinion; for the purposes of this survey, included respondents were those who considered themselves sufficiently knowledgeable to participate.

Many findings are unsurprising, such as researchers' high ranking of the importance of sleep duration (TST, TIB) and their interest in a device which can reliably differentiate sleep from inactivity and provide data security (Figure 3, Q14). Accurate measurement of sleep is a prerequisite for the use of any device in a research setting [7,32]. Moreover, ensuring privacy and data security is required by Institutional Review Boards (IRB) in order to conduct human research studies. However, the survey highlighted a few areas where CSTs could make improvements. Notably, estimations of sleep depth were considered less important to respondents than almost any other information about sleep

(Figure 3, Q14). It may be that researchers' disinterest in sleep depth estimations are related to the fact that the majority of CSTs do not measure sleep architecture like PSG (N1, N2, SWS, REM), but instead provide non-equivalent measures (e.g., light, deep, REM) with little documentation with regards to the scoring criteria [33-35]. It is possible that researchers would be more interested in sleep staging capabilities if CST systems reliably measured PSG-equivalent sleep architecture, followed a standardized scoring criterion, or if sleep depth estimations were shown to be robustly related to sleep health outcomes.

While deep sleep may be connotated with superior quality, sleep depth was listed separately from measures of sleep quality in this survey. Objective sleep quality, such as number of awakenings, wake after sleep onset (WASO), or sleep efficiency (SE), was ranked by respondents as the second most important information about sleep and subjective sleep quality, or the sleeper's personal satisfaction with their sleep, was ranked fourth. Both these measures have independently been shown to be important for health and performance outcomes [36-40], which may explain why they were highly ranked by sleep experts. The value of nap detection and the minimum time period required for sleep determination is indicated by the finding that 83% of respondents collect data related to napping and a combined 96% preferred naps to be defined as a sleep event less than or equal to 40 minutes (Figure 2, Q10-Q11). The number of times per day sleep occurs (Daily Sleep Intervals; DSI) was ranked ninth in importance (Figure 3, Q13). Nap information is important to researchers working in operational domains like shiftwork, healthcare, or aviation. In these safety-sensitive industries, strategic napping is recommended to prevent on-the-job fatigue [41-46]. Napping is also important as a health indicator across all research populations [47-49]. The impact of napping on downstream effects cannot be assessed if devices do not record them.

Respondents overwhelmingly indicated that periods of inactivity shorter than an hour could reliably be considered a nap (Figure 2, Q11). No respondents indicated that minimum nap detection should be greater than 60 minutes. Few studies have assessed the performance of CSTs for measuring naps and the specific parameters regarding minimum duration for scoring sleep are poorly communicated by CST companies [7]. A CST (the Zulu watch) that has been validated against PSG [50] automatically detects sleep episodes as short as 20 minutes using only accelerometry, indicating that algorithms can be developed to detect short naps. Automatic detection of short sleep periods and multiple sleep periods per day is one area in which developers could improve CSTs with regards to scientific relevancy.

The importance of battery life was highlighted by a number of survey items. Battery life was ranked as the most important factor which limited researchers' observation window (Q15), and extended battery life (defined as greater than/equal to 30 days) was the fourth highest ranked feature of a device which would facilitate data collection (Q14). Having to recharge or replace the battery frequently can result in periods of missing data during data collection. The majority of respondents (44%) preferred a continuous data collection window between 4-14 days long based on responses to Q12. Thirty-four percent (34%) preferred 15-30 days as an observation window. As shown in Fig. 4, economic value and demand were greater for hypothetical devices with longer battery life on the HPT. The findings suggest that extending battery life is another area in which CST developers could enhance competitive value in the scientific market.

Respondents' preferences were occasionally contradictory. For example, respondents preferred devices which measure sleep at the wrist (Q7), but believed the most appropriate method for determining sleep onset in real-world environments was by measuring brain activity in combination with motor activity and biometric data (Q9).   Brain activity cannot currently be measured at the wrist. The next most popular method of sleep/wake determination, which was selected by 31% of respondents, was by a combination of motor activity and biometrics. This is how many CSTs determine sleep, but is not the method used by actigraphy [15,51]. Only 7% of respondents preferred a standard

actigraphy method for sleep-wake determination, either motor activity alone or motor activity in combination with self-report (Q9). This finding would suggest that researchers actually prefer the sleep determination method used by CSTs to actigraphy.

Respondents also rated data security and remote data extraction closely with respect to desirable features (Q14). Demand was greatest for a hypothetical device which had options of either remote or wired data extraction, but lowest for a device that only featured remote data extraction. Remote data extraction would likely rely on wireless technology which utilizes an automatic sync function to upload data to a cloud-based or internet-enabled server. Ensuring data security through a system such as this is not impossible but presents more areas of vulnerability than a wired data transfer to an offline computer.

The behavioral economic analysis of demand curves for three hypothetical devices provides insight into decision making behavior about purchasing CSTs for research. HPTs are typically used to assess individual commodity demand (for example, demand for alcohol and drugs) and their reliability and validity have been well- documented [52,53]. The demand curves created here have demonstrated significantly lower and more elastic demand for Device B than the other two devices. The main features that distinguish the device with the lowest demand (Device B) and the device with the highest demand (Device C) were long battery life (4 days for Device B, 30 days for Device C), and the ability to choose between extracted data wirelessly via Bluetooth or through a wired download (USB). Long battery life was indicated as one of the most desired features for CSTs, and short battery life was also indicated as the most limiting for research studies. Device A featured longer battery life (30 days), and data that could be extracted only through a USB download. There was significantly more demand for Device A than for Device B despite the fact that Device B included more features, including sleep depth estimation which had been validated against PSG. This finding supports the ranking of sleep depth estimation of having low importance to sleep researchers from Figure 3. From these data, it is likely that demand for Devices A and C were largely driven by their longer battery life features and secure wired data extraction features.

The low importance of sleep depth scoring, and the high importance of short nap detection, data security, and battery life is at odds with the features of the majority of CST wearables on the market today. IBR is conducting a follow-up analysis to compare features of currently-available devices against expert preferences. The contrast between what scientists consider important, and what is being produced suggests a fundamental shortcoming of the market driven approach to the design of sleep-tracking wearables. The average consumer is not necessarily the best judge of what is important to measure about a physiological process like sleep despite their interest in the health science aspect of a device. This gap is also an opportunity; it is clear that the current technologies are capable of delivering scientifically-important features. A company that orients toward scientifically-important features would have a unique market advantage, perhaps not just within the academic research community but the larger consumer populace as well. The next step in this project is to determine whether the average consumer is willing to pay more for a device with sleep-tracking features that the scientific community considers important and to measure how scientific validation or endorsement impacts hypothetical demand for a sleep monitoring wearable in consumers who indicate an interest in monitoring sleep.

Because the actual global population of sleep researchers who have expertise in data collection outside the laboratory is unknown, the statistical power of this sample size (N=46) is indeterminate. Great efforts were taken to recruit as many real-world sleep experts as possible, and a variety of regions and research domains were ultimately represented (Figure 1). However, these findings are limited to a qualitative analysis. The survey was conducted in English, which may have led to skewed demographics. Another limitation is that respondents were not evenly distributed across research domains, precluding a comparison between domains or years' experience regarding CST

preferences. Respondents had, on average, 11 years' experience conducting sleep research, with 10 years being the most frequently-indicated length of experience. It is interesting to note that CSTs have only been around for 10-11 years themselves [1,5]. Respondents were not asked to indicate their age, level of education, or terminal degree, so it is possible that the consensus opinions reflect those of early-career researchers. The survey was purposely designed not to ask these questions under the assumption that non-academic researchers, technicians, and students have a valuable perspective on real-world data collection that should not be discounted because of their demographics or degree status. However, an interesting follow-up study would be to evaluate the impact of researcher status on CST preferences or the economic value of their scientific endorsement.

This investigation of CSTs for sleep research purposes differs with respect to previous analyses because it focuses on the desirability of features to a sleep scientist rather than focusing on the accuracy or validity of sleep measurement. The importance of validation testing of CSTs has been well-established in the literature [7,9-11,15], but even a validated device may not translate to researcher desirability. Four (4) respondents provided additional feedback about CSTs in the comments section at the end (see Appendix A). Each of these comments expresses a concern that the data are somehow invalid—either due to arbitrary units, changing algorithms, or improper recording of sleep. While beyond the scope of these analyses, an interesting follow-up survey would be to assess the importance of trust and transparency in the relationship between sleep researchers and CSTs.

Despite the growing conversation about the viability of CSTs for research, manufacturers may not be interested in increasing scientific accuracy in their devices unless doing so is expected to result in greater consumer sales. While the global sleep tracking device market is estimated to be worth up to $50 billion by 2027 [54], the monetary value of scientific endorsement of a product or specific scientifically-relevant design features has not been quantified. The next study in our project will address whether increasing the value of a CST product within the sleep research community, backed by independent validation and endorsement, results in an increase in value of the product for the general consumers interested in purchasing a sleep tracker.

5. Conclusions

Consensus opinion from this survey indicates that real-world sleep experts are looking for an accurate and reliable multi-sensor wrist-worn device that measures sleep on an epoch-by-epoch basis, detects sleep as short as 20 minutes, and has sufficient battery life to record data continuously for between 4-30 days. Real-world researchers are more interested in measures of sleep duration and quality than they are in sleep depth estimation, diagnostic information, circadian measures, or cognitive performance. The device should be able to remotely extract data but needs to provide data security, perhaps through a feature which allows researchers to turn off wireless capabilities. The features ranked most highly as important to real-world sleep researchers do not align with the most prominent features of currently-marketed wearables. These data provide context for further behavioral economics analyzes to determine differences in demand for scientifically-relevant CST device features between scientists and general consumers in order to inform IoT business and development decisions.

Author Contributions: "Conceptualization, J.K.D., L.P.S. and S.R.H.; methodology, J.K.D., L.P.S., J.C., and S.R.H.; formal analysis, J.K.D., L.P.S., J.C., and S.R.H.; data curation, J.C.; writing—original draft preparation, J.K.D., L.P.S. and S.R.H.; writing—review and editing, J.K.D., L.P.S., J.C., and S.R.H.; visualization, J.K.D. and L.P.S; supervision, S.R.H. All authors have read and agreed to the published version of the manuscript.

Appendix A: Comments

Comment 1: In my experience, the data/feedback on the device and/or accompanying phone app need to be immediately understandable and trusted by the user. I've seen compliance and buy-in drop off quickly when subjects don't feel like they are getting useful or trusted data on a daily or weekly basis. If the device even misses picking up just one or two sleep episodes or is really inaccurate (e.g., detecting sleep onset or offset >1h from reality, or not picking up a major awakening in the night) can break trust easily, especially in the beginning stages of a study. Then subjects think its just a junk device. So having a device with immediate buy-in is huge, especially for longer-term studies (>1 week) where subjects aren't necessarily interacting with researchers daily.

Comment 2: One of the concerns that we run into with consumer products are the security issues surrounding the pipeline for cloud-based data and the use of a participant's phone (app-based). There is also the issue that was not explicit in the questionnaire as to whether the raw data are available or only the consumer company provided summary (EBE or whole night). Given that the company algorithms can change (improve!?), this would disrupt a longitudinal study or a study conducted over multiple years.

Comment 3: Activity counts are arbitrary/meaningless units; they are undefined. Actigraphy is only reliable for measuring the daily timing of activity (wakefulness?). A defined, unit of measure for "activity count" needs to be developed/established....like mph, bpm. Variability in the sensitivity to same motion/movement among the same or different devices is too great. E.g., Place 2 or more actigraphy devices on your wrist (same Make: Model) and record simultaneously (1 min epoch) for a few hours or 24 hours. The activity count values for corresponding minutes are not similar.... not even close. If you don't have a unit of measure, how can you calculate/conclude anything?

Comment 4: Both actigraphy and CSTs are proxy measures of sleep and both require significantly more work to have any confidence at all that they can reliably and accurately distinguish sleep and wake.

References

1.   Lujan, M.R.; Perez-Pozuelo, I.; Grandner, M.A. Past, Present, and Future of Multisensory Wearable Technology to Monitor Sleep and Circadian Rhythms. *Frontiers in Digital Health* **2021**, 104.

2.   Kripke, D.F.; Mullaney, D.J.; Messin, S.; Wyborney, V.G. Wrist actigraphic measures of sleep and rhythms. *Electroencephalogr Clin Neurophysiol* **1978**, *44*, 674-676, doi:10.1016/0013-4694(78)90133-5.

3.   Chanson, S.T. Internet: History, impact, enabling technologies and potential problems. **1998**.

4.   Stippig, A.; Hubers, U.; Emerich, M. Apps in sleep medicine. *Sleep Breath* **2015**, *19*, 411-417, doi:10.1007/s11325-014-1009-6.

5.   Ko, P.-R.T.; Kientz, J.A.; Choe, E.K.; Kay, M.; Landis, C.A.; Watson, N.F. Consumer sleep technologies: a review of the landscape. *Journal of clinical sleep medicine* **2015**, *11*, 1455-1461.

6.   Kelly, J.M.; Strecker, R.E.; Bianchi, M.T. Recent developments in home sleep-monitoring devices. *ISRN Neurol* **2012**, *2012*, 768794, doi:10.5402/2012/768794.

7.   Depner, C.M.; Cheng, P.C.; Devine, J.K.; Khosla, S.; de Zambotti, M.; Robillard, R.; Vakulin, A.; Drummond, S.P.A. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep* **2020**, *43*, doi:10.1093/sleep/zsz254.

8.   Kolla, B.P.; Mansukhani, S.; Mansukhani, M.P. Consumer sleep tracking devices: a review of mechanisms, validity and utility. *Expert review of medical devices* **2016**, *13*, 497-506.

9.   Goldstein, C.A.; Depner, C. Miles to go before we sleep… a step toward transparent evaluation of consumer sleep tracking devices. **2021**.

10.   Menghini, L.; Cellini, N.; Goldstone, A.; Baker, F.C.; de Zambotti, M. A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. *Sleep* **2021**, *44*, zsaa170.

11.   de Zambotti, M.; Cellini, N.; Menghini, L.; Sarlo, M.; Baker, F.C. Sensors capabilities, performance, and use of consumer sleep technology. *Sleep medicine clinics* **2020**, *15*, 1-30.

12.   Khosla, S.; Deak, M.C.; Gault, D.; Goldstein, C.A.; Hwang, D.; Kwon, Y.; O'Hearn, D.; Schutte-Rodin, S.; Yurcheshen, M.; Rosen, I.M. Consumer sleep technology: an American Academy of Sleep Medicine position statement. *Journal of clinical sleep medicine* **2018**, *14*, 877-880.

13.   Peake, J.M.; Kerr, G.; Sullivan, J.P. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. *Frontiers in physiology* **2018**, *9*, 743.

14.   Roomkham, S.; Lovell, D.; Cheung, J.; Perrin, D. Promises and challenges in the use of consumer-grade devices for sleep monitoring. *IEEE reviews in biomedical engineering* **2018**, *11*, 53-67.

15.   Chinoy, E.D.; Cuellar, J.A.; Huwa, K.E.; Jameson, J.T.; Watson, C.H.; Bessman, S.C.; Hirsch, D.A.; Cooper, A.D.; Drummond, S.P.; Markwald, R.R. Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep* **2021**, *44*, zsaa291.

16.   Hursh, S.R.; Roma, P.G. Behavioral economics and empirical public policy. *J Exp Anal Behav* **2013**, *99*, 98-124, doi:10.1002/jeab.7.

17.   Hursh, S.R. Behavioral economics. *J Exp Anal Behav* **1984**, *42*, 435-452, doi:10.1901/jeab.1984.42-435.

18.   Hursh, S.R. Economic concepts for the analysis of behavior. *J Exp Anal Behav* **1980**, *34*, 219-238, doi:10.1901/jeab.1980.34-219.

19.   Devine, J.K.; Schwartz, L.P.; Hursh, S. What do researchers want in a consumer sleep technology? *Sleep* **2021**, *44*, doi:10.1093/sleep/zsab078.

20.   Christopoulos, D. Peer Esteem Snowballing: A methodology for expert surveys. In Proceedings of the Eurostat conference for new techniques and technologies for statistics, 2009; pp. 171-179.

21.   Singh, A.S.; Masuku, M.B. Sampling techniques & determination of sample size in applied statistics research: An overview. *International Journal of economics, commerce and management* **2014**, *2*, 1-22.

22.   Choe, E.K.; Consolvo, S.; Watson, N.F.; Kientz, J.A. Opportunities for computing technologies to support healthy sleep behaviors. In Proceedings of the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011; pp. 3053-3062.

23.   Newington, L.; Metcalfe, A. Factors influencing recruitment to research: qualitative study of the experiences and perceptions of research teams. *BMC medical research methodology* **2014**, *14*, 1-11.

24.   Plummer, M.L.; Chan, A.; Kohl, K.; Taylor, A.B.; Baltag, V.; Saewyc, E.; Ross, D.A. Results of a Global Survey of Experts to Categorize the Suitability of Interventions for Inclusion in School Health Services. *Journal of Adolescent Health* **2021**.

25.   Wildermann, N.E.; Gredzens, C.; Avens, L.; Barrios-Garrido, H.A.; Bell, I.; Blumenthal, J.; Bolten, A.B.; McNeill, J.B.; Casale, P.; Di Domenico, M. Informing research priorities for immature sea turtles through expert elicitation. *Endangered Species Research* **2018**, *37*, 55-76.

26. Dooley, K.J.; Corman, S.R.; McPhee, R.D. A knowledge directory for identifying experts and areas of expertise. *Human Systems Management* **2002**, *21*, 217-228.

27. Ghaljaie, F.; Naderifar, M.; Goli, H. Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in Development of Medical Education* **2017**, *14*.

28. Biernacki, P.; Waldorf, D. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research* **1981**, *10*, 141-163.

29. Stein, J.S.; Koffarnus, M.N.; Snider, S.E.; Quisenberry, A.J.; Bickel, W.K. Identification and management of nonsystematic purchase task data: Toward best practice. *Exp Clin Psychopharmacol* **2015**, *23*, 377-386, doi:10.1037/pha0000020.

30. Hursh, S.R.; Silberberg, A. Economic demand and essential value. *Psychol Rev* **2008**, *115*, 186-198, doi:10.1037/0033-295X.115.1.186.

31. Gilroy, S.P.; Kaplan, B.A.; Schwartz, L.P.; Reed, D.D.; Hursh, S.R. A zero-bounded model of operant demand. *J Exp Anal Behav* **2021**, *115*, 729-746, doi:10.1002/jeab.679.

32. De Zambotti, M.; Godino, J.G.; Baker, F.C.; Cheung, J.; Patrick, K.; Colrain, I.M. The boom in wearable technology: cause for alarm or just what is needed to better understand sleep? *Sleep* **2016**, *39*, 1761-1762.

33. De Zambotti, M.; Cellini, N.; Goldstone, A.; Colrain, I.M.; Baker, F.C. Wearable sleep technology in clinical and research settings. *Medicine and science in sports and exercise* **2019**, *51*, 1538.

34. Bhat, S.; Ferraris, A.; Gupta, D.; Mozafarian, M.; DeBari, V.A.; Gushway-Henry, N.; Gowda, S.P.; Polos, P.G.; Rubinstein, M.; Seidu, H.; et al. Is There a Clinical Role For Smartphone Sleep Apps? Comparison of Sleep Cycle Detection by a Smartphone Application to Polysomnography. *J Clin Sleep Med* **2015**, *11*, 709-715, doi:10.5664/jcsm.4840.

35. Liang, Z.; Chapa-Martell, M.A. Accuracy of Fitbit Wristbands in Measuring Sleep Stage Transitions and the Effect of User-Specific Factors. *JMIR Mhealth Uhealth* **2019**, *7*, e13384, doi:10.2196/13384.

36. Aziz, M.; Ali, S.S.; Das, S.; Younus, A.; Malik, R.; Latif, M.A.; Humayun, C.; Anugula, D.; Abbas, G.; Salami, J.; et al. Association of Subjective and Objective Sleep Duration as well as Sleep Quality with Non-Invasive Markers of Sub-Clinical Cardiovascular Disease (CVD): A Systematic Review. *J Atheroscler Thromb* **2017**, *24*, 208-226, doi:10.5551/jat.36194.

37. Bartolacci, C.; Scarpelli, S.; D'Atri, A.; Gorgoni, M.; Annarumma, L.; Cloos, C.; Giannini, A.M.; De Gennaro, L. The influence of sleep quality, vigilance, and sleepiness on driving-related cognitive abilities: A comparison between young and older adults. *Brain sciences* **2020**, *10*, 327.

38. Kwok, C.S.; Kontopantelis, E.; Kuligowski, G.; Gray, M.; Muhyaldeen, A.; Gale, C.P.; Peat, G.M.; Cleator, J.; Chew-Graham, C.; Loke, Y.K.; et al. Self-Reported Sleep Duration and Quality and Cardiovascular Disease and Mortality: A Dose-Response Meta-Analysis. *J Am Heart Assoc* **2018**, *7*, e008552, doi:10.1161/JAHA.118.008552.

39. Medic, G.; Wille, M.; Hemels, M.E. Short- and long-term health consequences of sleep disruption. *Nat Sci Sleep* **2017**, *9*, 151-161, doi:10.2147/NSS.S134864.

40. Zavecz, Z.; Nagy, T.; Galkó, A.; Nemeth, D.; Janacsek, K. The relationship between subjective sleep quality and cognitive performance in healthy young adults: Evidence from three empirical studies. *Scientific Reports* **2020**, *10*, 4855, doi:10.1038/s41598-020-61627-6.

41. Hartzler, B.M. Fatigue on the flight deck: the consequences of sleep loss and the benefits of napping. *Accident Analysis & Prevention* **2014**, *62*, 309-318.

42. Martin-Gill, C.; Barger, L.K.; Moore, C.G.; Higgins, J.S.; Teasley, E.M.; Weiss, P.M.; Condle, J.P.; Flickinger, K.L.; Coppler, P.J.; Sequeira, D.J.; et al. Effects of Napping During Shift Work on Sleepiness and Performance in Emergency Medical Services Personnel and Similar Shift Workers: A Systematic Review and Meta-Analysis. *Prehosp Emerg Care* **2018**, *22*, 47-57, doi:10.1080/10903127.2017.1376136.

43.   Sensors and Wearables: Using Biometric Data to Detect and Monitor Fatigue. *Work to Zero Webinar Series* **2021**, https://www.nsc.org/workplace/safety-topics/work-to-zero/resources.

44.   Schwartz, L.P.; Hursh, S.R.; Boyle, L.; Davis, J.E.; Smith, M.; Fitzgibbons, S.C. Fatigue in surgical residents an analysis of duty-hours and the effect of hypothetical naps on predicted performance. *Am J Surg* **2020**, doi:10.1016/j.amjsurg.2020.08.015.

45.   Alger, S.E.; Brager, A.J.; Capaldi, V.F. Challenging the stigma of workplace napping. *Sleep* **2019**, *42*, zsz097.

46.   Smith, S.S.; Kilby, S.; Jorgensen, G.; Douglas, J.A. Napping and nightshift work: Effects of a short nap on psychomotor vigilance and subjective sleepiness in health workers. *Sleep and Biological Rhythms* **2007**, *5*, 117-125.

47.   Devine, J.K.; Wolf, J.M. Integrating nap and night-time sleep into sleep patterns reveals differential links to health-relevant outcomes. *J Sleep Res* **2016**, *25*, 225-233, doi:10.1111/jsr.12369.

48.   Faraut, B.; Andrillon, T.; Vecchierini, M.-F.; Leger, D. Napping: a public health issue. From epidemiological to laboratory studies. *Sleep medicine reviews* **2017**, *35*, 85-100.

49.   Milner, C.E.; Cote, K.A. Benefits of napping in healthy adults: impact of nap length, time of day, age, and experience with napping. *Journal of sleep research* **2009**, *18*, 272-281.

50.   Devine, J.K.; Chinoy, E.D.; Markwald, R.R.; Schwartz, L.P.; Hursh, S.R. Validation of Zulu Watch against Polysomnography and Actigraphy for On-Wrist Sleep-Wake Determination and Sleep-Depth Estimation. *Sensors (Basel)* **2020**, *21*, doi:10.3390/s21010076.

51.   Sadeh, A. The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev* **2011**, *15*, 259-267, doi:10.1016/j.smrv.2010.10.001.

52.   Amlung, M.; MacKillop, J. Further evidence of close correspondence for alcohol demand decision making for hypothetical and incentivized rewards. *Behav Processes* **2015**, *113*, 187-191, doi:10.1016/j.beproc.2015.02.012.

53.   Few, L.R.; Acker, J.; Murphy, C.; MacKillop, J. Temporal stability of a cigarette purchase task. *Nicotine Tob Res* **2012**, *14*, 761-765, doi:10.1093/ntr/ntr222.

54.   *Global Smart Sleep Tracking Devices Sales Market Report 2021*; QYR-17567525; 05 March 2021 2021.

1.