

Feasibility and acceptability of using wearable physiological monitors with suicidal adolescent inpatients

Evan M. Kleiman,^{1,2,3} Alexander J. Millner,^{2,3} Victoria W. Joyce,³ Carol C. Nash,³ Ralph J. Buonopane,³ & Matthew K. Nock^{2,3}

1. Department of Psychology, Rutgers, The State University of New Jersey
2. Department of Psychology, Harvard University
3. Mental Health Research Program, Franciscan Hospital for Children

Acknowledgements:

Research reported in this publication was supported by the National Institute of Mental Health (R21MH115293; EMK) and the Chet and Will Griswold Suicide Prevention Fund (MKN).

Corresponding author:

Evan M. Kleiman
Department of Psychology
Rutgers, The State University of New Jersey
Tillett Hall, Room 627
53 Avenue E, Piscataway, NJ 08854
evan.kleiman@rutgers.edu

Abstract

Background: Wearable physiological monitoring devices enable the continuous measurement of human behavior and psychophysiology in the real world. Although such monitors are promising, their availability does not guarantee that participants will continuously wear and interact with them, especially during times of psychological distress.

Objectives: The goal of this study was to evaluate the feasibility and acceptability of using a wearable behavioral and physiological monitor, the Empatica E4, to continuously assess a group of suicidal adolescent inpatients.

Methods: Participants ($n = 50$ adolescent inpatients) were asked to wear an Empatica E4 on their wrist for the duration of their inpatient stay. In addition to assessing behavioral meta-data (e.g., hours worn per day) we also used qualitative interviews and self-report measures to assess participants' experience of wearing the monitor.

Results: Results supported the feasibility and acceptability of this approach. Participants wore the monitor for an average for 18 hours a day and reported that despite sometimes finding the monitor uncomfortable, they did not mind wearing it. Many of the participants noted that the part of the study they enjoyed most was contributing to scientific understanding, especially if it could help people like them in the future.

Conclusions: These findings provide promising support for using wearable monitors in clinical samples in future studies, especially if participants are invested in being part of a research study.

Introduction

Objective measures of physiological factors such as electrodermal activity (EDA) and heart rate (HR) have existed for over 100 years. Shortly after their development, researchers started using these tools in laboratory settings to examine the association between physiology and emotion [for historical reviews, see 1–3]. Recent technological advances have enabled the study of human behavior and psychophysiology outside of the laboratory, in the real world, using research-grade wrist-worn physiological monitors. These monitors enable the continuous, extended, real-world assessment of many of the same constructs once only possible to assess in lab settings over short assessment periods [4,5]. This offers great promise to improve our understanding, prediction, and prevention of factors related to psychological phenomena of interest. One area where technology may be particularly useful is suicidal thoughts and behaviors, which are highly prevalent among adolescents (15% of all adolescents each year seriously consider suicide [6]) and present increased risk for suicide death, which account for 8.5% of all deaths among adolescents and young adults globally [7]. For example, through other technology-based studies (i.e., studies using smartphone-based ecological momentary assessment), we know that suicidal thinking varies rapidly throughout the day and is associated with times of intense psychological distress (i.e., high-arousal negative affect) [8]. We also know from laboratory studies that physiological signals like EDA and HR map on to psychological distress [9–13]. We do not know, however, whether these EDA and HR are associated with suicidal thinking as it occurs in real time. Having this information would allow us to create interventions that can be triggered based on these physiological signals and delivered just-in-time as needed.

Public familiarity with wearables has increased in recent years with availability of consumer-grade monitors (e.g., Fitbit, Apple Watch). Unfortunately, these commercial monitors are limited for research use given that they far are less accurate than gold-standard laboratory-based monitors [14–16] and often do not include sensors to measure important psychophysiological variables like EDA. Compared to consumer-grade devices, scientific-grade wearable monitors are more accurate, but also more expensive, bulkier, and less user friendly than consumer-grade wearables. Thus, we cannot infer that high acceptability of consumer-grade wearables will translate to research-grade monitors. Accordingly, before such widescale research is possible, is important to assess the feasibility and acceptability of using these monitors in real-world research. The goal of the present study was to assess feasibility and acceptability of a commonly-used research-grade wearable physiological monitor (Empatica E4) to continually assess behavior and psychophysiology among a clinically severe group of adolescents – those hospitalized for suicidal thoughts and behaviors – over the course of their hospital stay. The Empatica E4 is a physiological monitor that is worn on the wrist like a watch and records several streams of physiological data, including EDA and HR (through an optical sensor), as well as temperature and movement (through an accelerometer).

There are three reasons it is important to test the feasibility and acceptability of these monitors within adolescent and psychiatric populations. First, prior studies testing the feasibility of wearable physiological monitors like the E4 have been among adult samples that are relatively psychologically healthy, such as people who suffer from migraines [17] and tourists visiting a new city [18]. These studies cannot tell us whether adolescents (who may find new technology

more acceptable or may be more self-conscious about the aesthetics of a wearable monitor) and/or those with more severe psychopathology (whose psychopathology may create competing demands for cognitive resources) would find these monitors acceptable. Second, nearly all studies have only collected data for a short period of time (e.g., 10 – 20 minutes [19,20]) and cannot tell us whether participants would find it acceptable to use the monitor over far longer periods (e.g., days, weeks, or months). Third, some newer wearable monitors are equipped with an event marker button that participants can use to report the experience of some psychological event/outcome of interest. Such a feature allows researchers to examine physiological data leading up to (and following) events of interest. However, there has been no exploration of whether it is feasible to ask adolescent participants to press an event marker button during times of transient psychological events (e.g., intense distress, severe suicidal thinking).

To our knowledge, only one study directly addressed feasibility/acceptability of the E4 [21]. This study found that a group of adults with schizophrenia and a control group were able to follow the instructions for using the E4 and rated it highly on a composite measure of acceptability. Although this information is useful, it leaves unaddressed several questions about the feasibility and acceptability of using such monitors with adolescents in acute psychiatric distress, and over longer periods of time. Thus, here we were interested in three questions relating to feasibility and acceptability that have not been addressed by prior studies.

First, we wanted to examine whether participants would wear the monitor the majority of each day over the course of multiple days. A monitor that can assess psychophysiology continuously throughout the day is only useful if participants are willing to wear the monitor over this period. Prior work has generally lasted only a few hours, and therefore could not examine whether participants are willing to continuously wear the monitor over days or weeks.

Second, we were interested in whether participants would interact with the monitor (i.e., use the self-initiated button press). Simply collecting behavioral and psychophysiological data does not enable researchers to document (and predict) behavioral or cognitive outcomes (e.g., psychological distress, suicidal thoughts, hallucinations). However, such a button is only useful if participants remember to press the button when they should and do not press the button when they should not. One prior study presents some data on whether adolescent participants could reliably press an event marker on a wearable monitor (actigraphy watch; [22]) each time they laid down to rest/sleep or got up from resting/sleeping. Although the authors of this study did not report actual compliance rates regarding the use of the event marker, they noted that “the event marker button was reliably used during the first few days but, afterward, some participants neglected to use it” (p. 555). The experience of transient psychological events such as intense distress may be more (or less) memorable than times before and after rest, and thus it is unknown whether participants will use the event marker at the same rate as this prior study. Third, we were interested in what participants liked (or disliked) about wearing the monitor. The only study to explicitly address participants’ opinions on the monitor [21] reported only that approximately 80% of their entire sample rated the monitor as “good” or “excellent” but did not report what participants specifically liked or disliked about the monitor.

Method

Participants

Data were drawn from the first 50 participants from a larger, ongoing study of suicidal adolescent inpatients assessing risk for harm to self or others using wearable ambulatory monitoring. Participants were recruited from a large urban inpatient psychiatry unit. Inclusion criteria for the study were: (a) admitted for risk of harm to self (e.g., severe suicidal ideation,

suicide attempt, non-suicidal self-injury), (b) being age 12-19, (c) and having at least one wrist with unbroken skin where the wristband could be placed.

Recruitment and data collection procedures

The study took place during participants' inpatient stay. Due to hospital policy, we were not able to compensate participants.

Consent.

We recruited and consented participants as close to hospital admission as possible. For potential participants under 18 years old, we first approached parents/guardians to get their written consent, and then approached the participant to get their written assent. We directly approached potential participants who were 18 years of age or older and received written consent. All study procedures were approved by the governing hospital and university IRBs.

Baseline measures and wearable monitor training.

After providing informed consent, participants completed a brief set of self-report measures as part of the larger study but not relevant to this manuscript and so not discussed further here (e.g., measures of emotion regulation, impulsiveness). Next participants completed a brief training session on how to properly wear and use the E4 and received a laminated one-page information sheet with the same instructions to serve as a personal reference.

Monitoring period.

For the duration of the inpatient stay, we asked participants to wear an E4 on their dominant wrist as often as possible (e.g., during the day, while sleeping) as long as the monitor was not at risk of getting wet (e.g., during showers). The E4 has an event marker button that can be used to "tag" events defined by the research team. We asked participants to press the marker button on the E4 whenever they felt distressed, which we defined as "Feeling so upset or angry

that you have an urge to hurt yourself or someone else or to break something.” We made sure that participants were aware that no one was actively monitoring when they pressed the button (i.e., pressing the button would not signal the clinical team to come help them and we would not share the study information with the clinical team). Each day, a study staff member (during the work week) or a clinical staff member (during the weekend) switched each participant’s E4 for a fully charged monitor.

Daily check-in surveys.

Each weekday, a study staff member approached the participant to conduct a brief check-in about any problems with the E4 that occurred since the last check-in. Staff members also assessed: (1) whether the participant recalled missing any occasions when they believe they should have pressed the button but did not and (2) whether the participant accidentally pressed the button. If a participant recalled missing a button press or accidentally pressed the button, we assessed when and why this occurred.

Discharge.

Shortly before their discharge from the hospital, participants completed two sets of open-ended questions aimed at assessing their experiences wearing the E4. First, they completed a 12-item questionnaire regarding satisfaction with the E4, modelled after other measures of comfort with wearable devices, specifically the Wearable Computer Comfort Rating Scale developed by Knight & Baber [23]. Items on this measure assessed: (1) concerns about appearance when wearing the device (e.g., “I felt anxious wearing the device”), (2) the physical feel of the device (e.g., “The device was uncomfortable to wear”), (3) whether the device affected movement (e.g., “the device made it hard to sleep at night”), and (4) general worries about taking care of the device (e.g., “I worried about taking care of the device”). All items were on a 0 (never) to 10 (all

the time) scale. Second, participants completed four open-ended qualitative questions: (1) “What did you like MOST about wearing the device?”, (2) “What did you like LEAST about wearing the device?”, (3) “How did you feel when wearing the device?” and (4) “Is there anything you would change about the device?”

The wearable monitor

Overview.

The Empatica E4 is a research grade wrist-worn behavioral and psychophysiological monitor (<https://www.empatica.com/en-eu/research/e4/>). Its case is 44mm long (~1.73 inches), 40mm wide, and 16mm deep. This means that it is larger than commercially available wearable monitors (e.g., the Fitbit Charge HR 2 is 22.86mm long, and 12.7mm wide, and 11.0mm deep). It has four main sensors: (1) an LED-based photoplethysmograph (PPG) used to derive heart rate from blood flow, (2) a pair of silver-plated EDA/skin conductance sensors, (3) a three-axis accelerometer, and (4) an infrared thermopile used to determine temperature. The E4 collects these data in real-time and stores them on the onboard flash memory (which can hold ~60 hours of data, at 1MB per hour). The E4 is then connected to a computer via a USB cradle and synced to a secure cloud server through the “E4 Connect” software. The E4 has a 250mAh battery (lasting ~36 hours) that charges via the syncing cradle. The E4 also offers Bluetooth streaming which can transmit data to the cloud using a compatible smartphone as a gateway. We did not use this option because it would not have been feasible for long-term use due to increased battery consumption as a result of using the Bluetooth radio and limitations in mobility (participants would need to be <30 feet from a smartphone at any time).

Placement of E4 on dominant wrist.

The larger goal of this study is to test whether the physiological and behavioral data collected from the E4 can accurately predict episodes of self- and other-directed violence. Thus, we made the decision about which wrist to wear the E4 based on our goal of collecting this data stream. There is currently debate in the field about which wrist is optimal for assessing EDA. A body of work suggests that high arousal negative emotion (e.g., distress) can be detected through examining asymmetry between the left and right sides of the body [24], with more pronounced signals coming from the dominant side. Given that wearing two monitors at once would be cumbersome for participants, we elected to have participants wear the monitor on their dominant wrist, which would likely provide the most pronounced changes in EDA in response to distress. This is also in line with the manufacturer's recommendation [25]. It is important to note, however, that the dominant wrist likely produces noisier data (due to motion artifacts) than the non-dominant wrist. Thus, researchers should examine the tradeoff between signal and noise when choosing which wrist to use in their own work.

Logistics of charging and syncing E4.

Each participant was assigned two wristbands, but only wore one wristband at a time. While the participant wore one wristband, the other was placed on a dock to charge and sync data. We assigned participants two wristbands because charging and syncing the E4 can, in some cases, take over an hour. Thus, having just one E4 per participant would have meant that we would lose over an hour of data per day while the monitor was charging. Since the E4 software can only sync two E4s at a time, it would have been cumbersome to have only one monitor per participant, since it would have required study staff to rotate the E4 on the syncing cradle throughout the day. To accommodate a larger number of participants (we had up to 10 simultaneously), we constructed a charging and syncing station that we could place the monitors

on when they were not being used, which was kept in a research office adjacent to the inpatient unit. The monitors would charge simultaneously and sync consecutively (i.e., when one E4 was finished syncing, the computer moved on to the next one).

Analytic strategy

How often will participants wear the monitors?

To answer this question, we calculated two features from the E4's recording length metadata: (1) number of days that each participant wore the monitor for at least some amount of time and (2) number of hours during each day that participants wore the monitor. From these features, we calculated the intra-class correlation (*ICC*) of hours per day wearing the monitor. This allowed us to examine how much of the variability in time wearing the band was due to between-person differences (i.e., whether some participants consistently tended to wear the band for more or less time than other participants) versus within-person differences (i.e., whether participants wore the band a lot more or less frequently on some days than other days). We also compared these features by attrition status (i.e., between those who did and did not drop out of the study).

Do participants correctly use the event marker?

To answer this question, we calculated three sets of features: (1) number of button presses, extracted from the raw data from the tags.csv file on the E4, (2) the number of missed button presses (and the reasons for missed press) extracted from the daily participant surveys, and (3) the number of accidental button presses (and reasons for the accidental press) extracted from the daily surveys. We also examined the *ICC* for number of button presses each day, again to determine whether some patients tended to press the button more often than others or if patients pressed the buttons more on some days than others.

Do participants like wearing the monitor?

To answer this question with the quantitative data, we calculated the means and standard deviations for each item on the comfort assessment measure. We explored the association between each item on the measure and length of time wearing the band each day. We also compared responses on the measure between those who did and did not drop out of the study. Both sets of comparisons were corrected for multiple comparisons using a Bonferroni Correction. We used a three-step process to answer this question using qualitative data. First, the first author first did a line-by-line read-through of the responses to the four qualitative questions and developed a codebook (described below). Second, the first, second, and last author independently coded the responses based on this codebook. Third, we resolved any discrepancies to reach a final consensus. As with the quantitative data, we also examined associations between the qualitative data (i.e., differences by whether participants did or did not endorse a qualitative category) and length of time wearing the monitor each day (using a t-test) and in drop-out status (using a chi-square test).

Results

Descriptive statistics

Participants' ages ranged from 12.5 to 18.6 years ($M = 16.3$ years, $SD = 1.6$ years), 78% of the sample was female and 92% of the sample was White (4% were Asian and the remainder indicated that they identified with another race). Participants together provided 487 total days of data ($M = 9.74$ days per participant, $SD = 13.81$ days, range = 1 – 76 days). The average length of stay in the hospital (i.e., day of intake to day of discharge) was 10.7 days ($SD = 13.86$ days, range = 1 – 77 days). There were no demographic differences by age or sex (there were too few cases to examine racial differences) on any key study variables including hours wearing the

monitor per day (age: $r = .05, p = .768$; sex: $t = 0.13, p = .896$); number of button presses (age: $r = .10, p = .579$; sex: $t = 1.26, p = .221$), and dropout status (age: $t = 0.09, p = .930$; sex: $\chi^2 = 0.78, p = .376$).

Do participants wear the monitor?

Participants wore the E4 at some point during the day (i.e., any non-zero amount of data for the day) for 464 out of a possible 487 total study days (95.3% of all days; see **Figure 1**, top panel). The majority (15/23; 65%) of days with no data were from one participant (#10 in the figures), although this participant was in the study for the longest amount of time (76 days). As can be seen in the bottom panel of **Figure 1**, there was considerable variability in the average number of hours each participant wore the band each day. On days when participants wore the E4, they did so for an average of 18.3 hours ($SD = 6.3$). Excluding the first and last day of the study, when participants would not have been able to wear the monitor for 24 hours, participants wore the monitor for an average of 20.3 hours ($SD = 5.3$). There was no association between day in study and hours worn ($b = -0.10, p = .465$), suggesting that participants did not wear the monitor any more or any less as their time in the study increased. When visually inspecting the scatterplot of this association, there was no clear non-linear effect, suggesting that participants did not likely wear the band more towards the middle of the study than at the ends. Regarding variability in hours worn per day, there was more within-person (i.e., day-to-day) variability in hours worn per day than there was between-person variability ($ICC = .31, 95\% CI = .21 - .43$). This means that one-third of the variability was due to some participants tending to wear the E4 longer than others whereas two-thirds of the variation in the amount of time wearing the monitor was accounted for by day-to-day variation within people.

Dropout. Seven participants discontinued wearing the monitor before they were discharged. As expected, those who dropped out of the study participated for significantly fewer days ($M = 2.1, SD = 1.4$) than those who continued the study until they were about to be discharged from the hospital ($M = 11.2, SD = 14.7; t = 3.92, p < .001$); and for significantly fewer hours each day when they were in the study ($M = 7.4, SD = 6.3$) than those who continued in the study ($M = 18.7, SD = 6.5; t = 6.61, p < .001$).

Do participants correctly use the event marker?

There were 2159 total button presses (i.e., uses of the event marker) recorded during the study, which occurred during 435 out of the 464 days (93.8%) during which participants wore the monitor. Participants pressed the button on average 4.9 times per day ($SD = 9.3$, range = 0 – 140). The top panel of **Figure 2** shows a plot of the daily number of button presses for each participant during each day they were in the study.

As can be seen in the bottom panel of **Figure 2**, there was considerable variability in the average number of presses per participant per day ranging from 0.8 presses/day to 77.7 presses/day. In contrast to the amount of time wearing the monitor, there was more between-person variability in average number of button presses per day than there was within-person variability ($ICC = .63, 95\% CI = .52 - .73$). This means that most of the variability in button presses was between-individuals, and each participant tended to stay near their own average throughout the study. As would be expected, when looking within each day regardless of subject, there was a small, but significant, association between hours worn per day and number of button presses ($B = 0.09, 95\% CI = 0.01 - 0.16, p = .040$). When examining the association between day in study and number of times pressing the button, there was a small but statistically significant

association ($B = -0.10$, 95% $CI = -0.15 - -0.05$, $p < .001$) suggesting that as participants were in the study for a longer period of time, they pressed the button slightly less.

Missed presses.

There were 40 total missed button presses reported during the study. Seventeen of these missed presses came from one participant. The remaining 23 missed presses (57.5% of all presses) came from 14 participants ($M = 1.7$ presses/participant, $SD = 1.3$, range = 1 – 6). Regarding reasons for missed presses: nearly all of the missed presses (32/40; 80% of presses) were due to the participant forgetting to press the button. The remainder of missed presses were because they were not wearing the monitor or the battery had died (4/40; 10% of presses), or due to an inability to pick a specific point when feeling distressed (2/40; 5% of presses) or misunderstanding that they were supposed to press the button (2/40; 5% of presses).

Accidental presses.

There were 10 accidental button presses reported during the study (0.46% of all presses). No participant had more than one accidental button press. The reasons participants accidentally pressed the button fell into two broad categories: (1) because another patient pressed their button (6/10 presses) and (2) the result of trying to turn off the monitor by holding down the button, but accidentally releasing it too early (3/10). For the one remaining accidental press, the patient did not recall the circumstances.

Do participants like wearing the monitor?

Quantitative measures.

A summary of responses to the quantitative questions about device comfort is shown in **Table 1** and a visualization of the distribution of responses is shown in **Figure 3**. On average, across nearly all items assessing comfort, participants tended to rate their discomfort wearing the

monitor below five out of 10 (with 10 meaning more discomfort). As can be seen in the middle columns of **Table 1**, there was a significant negative correlation, between hours worn and ratings of how uncomfortable the monitor was, such that the more participants rated the monitor as uncomfortable, the fewer hours they wore it. No other correlations were significant after correcting for multiple comparisons. As can be seen in the rightmost columns of **Table 1**, after correcting for multiple comparisons, there were only two significant findings: those who dropped out were significantly more likely to rate the device as uncomfortable and more likely to note that they could feel the device. The qualitative data illustrated these quantitative findings well. One participant who dropped out said “I really wanted to keep it on so I could help but it was too uncomfortable” and another said “I could feel the silver [EDA electrodes] rubbing.”

Table 1. Quantitative assessment of wearable monitor comfort

Item	Descriptive		Correlation with time worn		Did not drop out		Attrition status Dropped out		Comparison	
	<i>M</i>	<i>SD</i>	<i>r</i>	<i>p</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
I was worried about how I looked when I wore the device.	1.87	2.26	-0.14	.450	1.93	2.29	1.33	2.31	0.42	.706
I felt tense or on edge because I was wearing the device.	1.59	2.05	-0.04	.817	1.66	2.13	1.00	1.00	0.94	.399
I felt strange wearing the device.	2.91	2.67	-0.19	.265	2.65	2.35	5.00	4.40	-1.05	.365
I felt anxious wearing the device.	1.63	2.22	0.07	.730	1.78	2.29	0.33	0.58	2.61	.022
The device was uncomfortable to wear.	5.23	2.90	-0.49	.001	4.53	2.60	8.86	1.07	-7.30	.001
I could feel the device on my wrist.	7.28	2.62	-0.26	.085	6.87	2.63	9.57	0.79	-5.24	.001
The device interfered with my movement.	2.35	2.55	-0.13	.467	2.04	2.13	3.83	3.87	-1.10	.315
The device made it hard to sleep at night.	3.53	3.50	-0.07	.680	3.50	3.38	3.67	4.46	-0.09	.933
The device interfered with parts of my day.	3.15	3.20	-0.31	.078	2.67	2.95	5.33	3.67	-1.66	.143
I worried about taking care of the device.	3.51	3.28	0.02	.900	3.59	3.20	3.00	4.12	0.31	.773
I liked wearing the device.	3.60	2.96	0.30	.049	3.66	2.97	3.29	3.15	0.29	.779
Other people ask about the device.	4.46	3.27	0.33	.038	4.69	3.14	2.80	4.09	1.00	.368

Note. *M* = mean, *SD* = standard deviation. **Bold** = significant after Bonferroni correction (.05/24 = .002).

Qualitative measures.

The initial read-through of the qualitative responses yielded 13 codes across three categories: complaints about the monitor (7 codes), positive/neutral statements about the monitor (3 codes), and positive states about the study itself (3 codes). The results of the qualitative analyses are shown in the middle columns of **Table 2**. Our initial reliability across all codes was acceptable ($K = .77$, $SD = .20$, range = .041 - 1.00). The three raters were able to come to a consensus on all of the discrepancies. The rightmost column of **Table 2** shows the frequency of which each code was endorsed as well as example statements from each code.

The most commonly-endorsed codes were “feeling OK when wearing the monitor” (74.5% of the sample) and “liked helping in a research study” (59.6%). Interestingly, although the study only asked participants to monitor and express their distress to the extent needed to remember to press the button, some participants reported that they liked the monitoring because it allowed them to express distress (14.89%) or become more aware of distress (6.38%). On balance, nearly half of participants reported some discomfort while wearing the device (46.8%), with most common complaint that the device was too large (36.2%). As can be seen in **Table 3**, there were no significant differences between those who did or did not endorse any of the qualitative categories on hours worn per day or attrition status.

Table 2. Results of qualitative analyses

	Example	K	Z	% Endorsed
Complaints about monitor				
Discomfort/uncomfortable in general	<i>It felt extremely uncomfortable.</i>	0.68	8.08	46.81%
The monitor was too big/clunky	<i>It was bulky and inconvenient.</i>	0.90	10.74	36.17%
It should have a clock	<i>I would probably make it have a clock.</i>	0.92	10.89	23.40%
The material was uncomfortable	<i>I'd make the wristband out of a thinner material.</i>	0.41	4.83	14.89%
Discomfort sleeping/at night	<i>Was uncomfortable during sleep.</i>	0.92	10.93	10.64%
EDA sensors were uncomfortable	<i>Circular sensors are too big and rub too much.</i>	1.00	11.87	6.38%
The monitor does not look good	<i>It could be a little more sleek and comfortable.</i>	0.59	6.95	2.13%
Positive/neutral statements about device				
Felt OK when wearing monitor	<i>No different than normal.</i>	0.47	5.62	74.47%
Could tolerate negatives	<i>It isn't very comfortable but I managed.</i>	0.58	6.84	10.64%
Something positive about monitor's looks	<i>It looked cool.</i>	0.76	9.05	10.64%
Positive about paradigm/study				
Liked helping in a research study	<i>I felt I was helping.</i>	0.86	10.17	59.57%
Liked expressing distress	<i>I could press the button when I was in distress.</i>	0.94	11.15	14.89%
Helped become aware of distress	<i>I was able to be more alert and attentive to when I was having a hard time.</i>	1.00	11.87	6.38%

Note: *K* = Kappa from initial coding round.

Table 3. Differences in daily hours worn and attrition status by qualitative category endorsement

Category	Hours Worn						Attrition			
	<u>Endorsed</u>		<u>Did not endorse</u>		<i>t</i>	<i>p</i>	Did not drop out	Dropped out	<i>X</i> ²	<i>p</i>
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>							
Discomfort/uncomfortable in general	14.05	6.98	15.96	5.49	0.88	.389	45.0%	57.1%	0.35	.553
Discomfort sleeping/at night	16.19	5.25	15.01	6.32	-0.42	.698	12.5%	0.0%	0.98	.322
EDA sensors were uncomfortable	11.28	9.55	15.5	5.83	0.75	.525	5.0%	14.3%	0.86	.354
The monitor was too big/clunky	16.37	4.73	14.41	6.85	-1.00	.325	37.5%	28.6%	0.21	.650
The material was uncomfortable.	11.03	8.13	15.83	5.64	1.27	.264	12.5%	28.6%	1.21	.271
It should have a clock	15.99	5.40	14.85	6.46	-0.52	.610	22.5%	28.6%	0.12	.726
Felt OK when wearing monitor	15.70	5.49	13.75	7.70	-0.73	.479	77.5%	57.1%	1.30	.254
Could tolerate negatives	17.90	1.83	14.79	6.43	-2.11	.051	12.5%	0.0%	0.98	.322
Something positive about monitor 's looks	14.07	4.40	15.28	6.38	0.49	.647	10.0%	14.3%	0.12	.734
Liked expressing distress	12.13	8.75	15.64	5.65	0.87	.428	12.5%	28.6%	1.21	.271
Liked helping in a research study	16.32	5.46	13.14	6.93	-1.41	.172	57.5%	71.4%	0.48	.488

Note: Codes for “device does not look good” and “helped become aware of distress” were not included in these analyses due to low frequency of endorsement.

Discussion

This study examined and found support for the feasibility and acceptability of using wearable behavioral and psychophysiological monitors to continuously collect objective data from adolescents with clinically severe psychiatric problems. There are several key findings from this study. First, participants were compliant with wearing the monitor, doing so on average more than 18 hours per day. Second, participants were compliant with instructions to use the event marker when distressed and were able to do so independently and without prompting. Third, participants found wearing the monitor to be acceptable and liked wearing them because it was part of a research study (e.g., helping researchers understand psychiatric conditions that might someday be used to help others in a similar position). We discuss these main findings in greater detail below.

Do participants wear the monitor?

The first aim of the study was to assess how often participants wore the monitor. We found that participants wore the E4 at some point nearly every day and did so, on average, more than 18 hours a day. We also found that the variability in hours worn per day has more to do with daily-level factors than person-level factors. These data reflect that it may not be that there are certain types of participants that wear wristband more than others, but rather certain days where any given participant is more likely to wear the wristband than other days. Accordingly, future research could focus on reasons for this day-to-day variability to optimize participant compliance (e.g., by finding factors that can identify the types of days where participants are more or less likely to wear the wristband). Overall, these data suggest that it is possible to use wearable monitors to collect continuous, objective data from clinically severe adolescents as they navigate

their daily lives, and as such, to collect exponentially larger and more ecologically valid data than what has been possible in laboratory studies where recordings might be for a few hours at most while performing benign experimental tasks. This opens up myriad possibilities for better understanding the phenomenology and prediction of a range of clinical outcomes such as depression, anxiety, psychotic experiences, and suicidal and violent thoughts/behaviors.

Do participants correctly use the event marker?

The second aim of the study was to assess whether participants interacted with the monitor. We asked participants to press the event marker on the monitor when they were feeling distressed and found that they were generally compliant in doing so. Participants rarely accidentally pressed the button (approximately 1 in 216 button presses was an accident) and rarely forgot to press the button (participants forgot to press the button approximately 1 out of every 59 times they should have). This is particularly impressive given the circumstances under which we asked participants to press the button. Being on an inpatient unit can be unfamiliar and stressful and being highly distressed (which is when they were asked to press the button) is an affect state which may be particularly difficult to self-monitor and respond to. Indeed, the basis for this larger study these data are drawn from is to attempt to identify distress using the E4's sensors so that interventions could be developed that do not rely on patients needing to monitor for distress. Regarding the frequency of the button presses each day, we found that although there was day-to-day variability in the times each participant pressed the button each day, there was more variability from participant-to-participant, suggesting that some participants tended to use the event marker more than others. It is unclear whether this is best explained by between-participant variability in conscientiousness in pressing the button vs. in the likelihood of actually experiencing distress. Future work examining the correspondence between button presses and

actual physiological arousal (i.e., as an objective indicator of distress) will help to address this question. We also found that as the number of days in the study increased, the number of times pressing the button each day slightly decreased. This could be a marker of an effective treatment (i.e., participants were feeling less distressed) or a marker of study fatigue. Given how infrequently participants did not press the button when they should have, it is probably more likely that this is a marker of treatment efficacy rather than study fatigue.

Do participants like wearing the monitor?

The third aim of the study was to assess both quantitatively and qualitatively what participants liked and did not like about participating in a study using wearable physiological monitors. We found that participants tended to report that the monitor did not interfere much with their movement and that they “felt OK” when wearing the monitor. Many participants reported that one of the aspects they liked the most about the study was being in a study where they could contribute to knowledge acquisition and/or help others like them. On balance, nearly half of the sample noted that the monitor was generally uncomfortable, with over a third of participants noting the monitor was too bulky. It is also notable that more than 15% of participants said that they liked how the monitor helped them be more aware of their distress, even though this was not an explicit goal of the study. This is interesting because it may suggest the viability of interventions using these monitor to help participants self-monitor and manage their own distress in a more explicit manner.

Taken together, these findings suggest that doing research with these wearable monitors is feasible, especially when participants are motivated to be in a study. These findings also suggest, however that the feasibility of using wearable physiological monitors for research may not translate to applied settings where the incentive of being in a research study is not there to

motivate compliance. In these cases, there would be fewer incentives to balance out the negatives about the size of the monitor. Thus, more comfortable monitors may be required to obtain similar compliance to what we found in the present study. For example, the Empatica Embrace is a currently-available consumer monitor that is approximately 30% smaller than the E4 (30mm x 30mm x 10mm, 44mm x 40mm x 16mm, respectively). Although not explicitly designed for research use, it can be used in some research applications (e.g., studies concerned with movement and skin conductance, but not heart rate since it does not have a PPG sensor). Moreover, it is reasonable to expect that smaller research-grade monitors will be available from a variety of manufactures in the near future.

Limitations, strengths, and future directions

These findings should be interpreted in the context of three key limitations. First, although we assessed the *amount* of data received, we did not assess the *quality* of data received. Not all data collected from any wearable monitor are useable (e.g., due to motion artifacts). Evaluating the quality of the data obtained in such studies is an important step for future research, and this is something that we will be undertaking in our ongoing work in this area. Second, we used recording length as a proxy for time wearing the E4. If the E4 is turned on, it records even if it is not being worn. Thus, it is possible that the data reported here on time that the monitor was worn may be overestimated. There are two reasons why the probability of time worn overestimations is quite low. One is that participants were told to turn off the E4 whenever it was not being worn. Another is that data were collected on a psychiatric inpatient unit where unit staff observes patients at least once every five minutes. If unit staff saw an unworn E4 that was powered on, they would turn off the monitor. A third limitation, related to the prior point, is that these data were collected on an inpatient unit where participants were continually monitored

by clinical staff and visited nearly daily by research staff, possibly making them more likely to wear the monitors. Thus, findings from this study may not generalize to other settings where there is less intensive adult supervision (e.g., schools, home) or to other samples (e.g., adults). Testing the generality of these results in other settings and samples is another important future step for research in this area. Finally, although our study provided a rich description of how often participants wore the monitor and interacted with it, the study was unable in some cases to provide explanations of why participants wore a device more on some days than others. Like many studies whose goal is description, future studies should explore possible explanations for the phenomena described in this study.

In conclusion, the clearest implication from this study is that it is feasible to conduct research where participants wear physiological monitors for an extended period (i.e., days or weeks). This implication is in line with other studies of samples of adults with less severe psychopathology [21]. The study's findings are important because studies that use wearable behavioral and psychophysiological monitors over an extended period have great promise to help researchers understand how constructs of interest to psychological scientists operate in everyday life. This is especially true for studies that combine these data streams with other streams of data such as medical records, passive smartphone sensing, and ecological momentary assessment. Our ability to do studies like this is only just beginning and it is undoubtable that wearable technology will become even more advanced in the coming years, making studies even more feasible in the future.

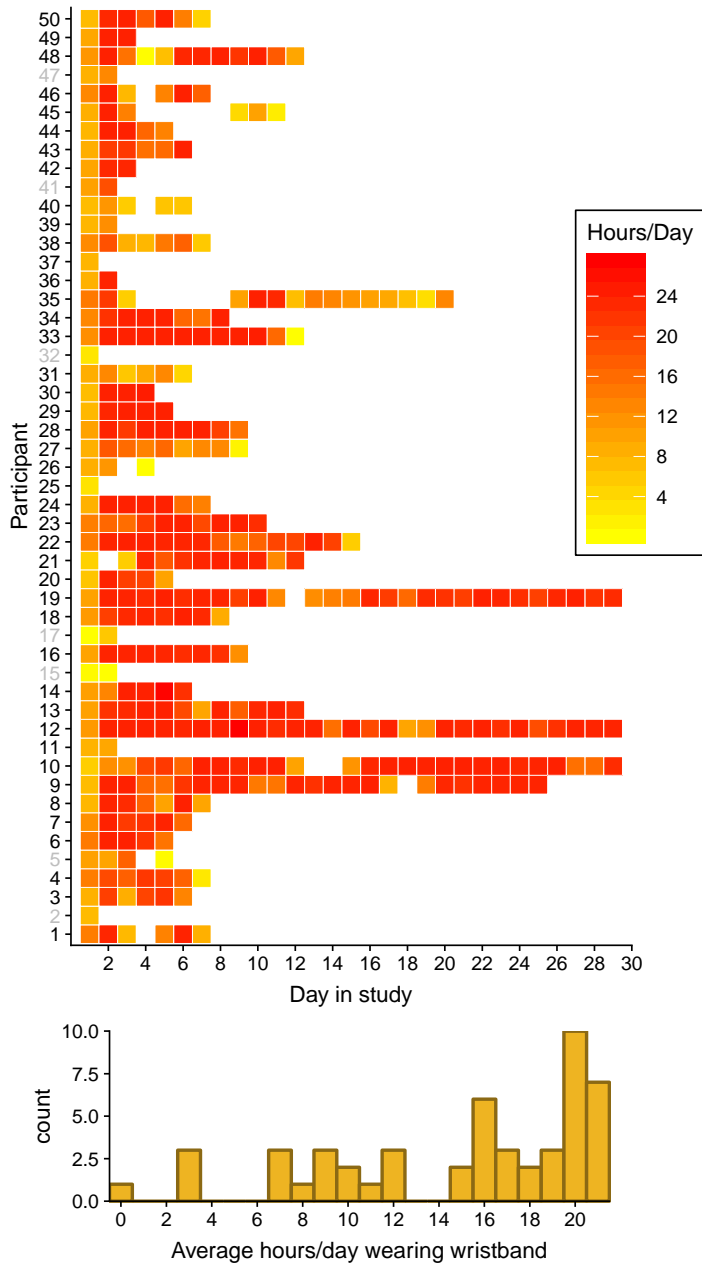
References

1. AlGhatrif M, Lindsay J. A brief review: history to understand fundamentals of electrocardiography. *J Community Hosp Intern Med Perspect* [Internet] 2012 Apr 30 [cited 2017 Jan 26];2(1). PMID:23882360
2. Fye WB. A history of the origin, evolution, and impact of electrocardiography. *Am J Cardiol* 1994 May 15;73(13):937–949. PMID:8184849
3. Neumann E, Blanton R. The early history of electrodermal research. *Psychophysiology* 1970;6(4):453–475. [doi: 10.1111/j.1469-8986.1970.tb01755.x]
4. Dolgin E. Technology: Dressed to detect. *Nature* 2014 Jul;511(7508):S16–S17. [doi: 10.1038/511S16a]
5. Poh MZ, Swenson NC, Picard RW. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Trans Biomed Eng* 2010 May;57(5):1243–1252. [doi: 10.1109/TBME.2009.2038487]
6. Cash SJ, Bridge JA. Epidemiology of youth suicide and suicidal behavior. *Curr Opin Pediatr* 2009 Oct;21(5):613–619. PMID:19644372
7. Cha CB, Franz PJ, Guzmán EM, Glenn CR, Kleiman EM, Nock MK. Annual Research Review: Suicide among youth – epidemiology, (potential) etiology, and treatment. *J Child Psychol Psychiatry* 2018;59(4):460–482. [doi: 10.1111/jcpp.12831]
8. Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, Nock MK. Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies. *J Abnorm Psychol* 2017 Aug;126(6):726–738. [doi: 10.1037/abn0000273]
9. Lensvelt-Mulders G, Hettema J. Genetic analysis of autonomic reactivity to psychologically stressful situations. *Biol Psychol* 2001 Sep 1;58(1):25–40. [doi: 10.1016/S0301-0511(01)00099-0]
10. Reinhardt T, Schmahl C, Wüst S, Bohus M. Salivary cortisol, heart rate, electrodermal activity and subjective stress responses to the Mannheim Multicomponent Stress Test (MMST). *Psychiatry Res* 2012 Jun 30;198(1):106–111. [doi: 10.1016/j.psychres.2011.12.009]
11. Lazarus RS, Speisman JC, Mordkoff AM, Davison LA. A laboratory study of psychological stress produced by a motion picture film. *Psychol Monogr Gen Appl* 1962;76(34):1–35. [doi: 10.1037/h0093861]

12. Garralda ME, Connell J, Taylor DC. Psychophysiological anomalies in children with emotional and conduct disorders. *Psychol Med* 1991 Nov;21(04):947. [doi: 10.1017/S0033291700029937]
13. Nock MK, Mendes WB. Physiological arousal, distress tolerance, and social problem-solving deficits among adolescent self-injurers. *J Consult Clin Psychol* 2008 Feb;76(1):28–38. PMID:18229980
14. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. *PLOS ONE* 2018 Feb 28;13(2):e0192691. [doi: 10.1371/journal.pone.0192691]
15. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep* 2015 Aug 1;38(8):1323–1330. PMID:26118555
16. Shoval N, Schvimer Y, Tamir M. Real-time measurement of tourists' objective and subjective emotions in time and space. *J Travel Res* 2018 Jan 1;57(1):3–16. [doi: 10.1177/0047287517691155]
17. Siirtola P, Koskimäki H, Mönttinen H, Röning J. Using sleep time data from wearable sensors for early detection of migraine attacks. *Sensors [Internet]* 2018 Apr 28 [cited 2018 Jun 30];18(5). PMID:29710791
18. Shoval N, Schvimer Y, Tamir M. Tracking technologies and urban analysis: Adding the emotional dimension. *Cities* 2018 Feb 1;72:34–42. [doi: 10.1016/j.cities.2017.08.005]
19. Corino VDA, Laureanti R, Ferranti L, Scarpini G, Lombardi F, Mainardi LT. Detection of atrial fibrillation episodes using a wristband device. *Physiol Meas* 2017;38(5):787. [doi: 10.1088/1361-6579/aa5dd7]
20. Spann CA, Schaeffer J, Siemens G. Expanding the scope of learning analytics data: Preliminary findings on attention and self-regulation using wearable technology. *Proc Seventh Int Learn Anal Knowl Conf [Internet]* New York, NY, USA: ACM; 2017 [cited 2018 Jun 30]. p. 203–207. [doi: 10.1145/3027385.3027427]
21. Cella M, Okruszek Ł, Lawrence M, Zarlenga V, He Z, Wykes T. Using wearable technology to detect the autonomic signature of illness severity in schizophrenia. *Schizophr Res* 2018 May;195:537–542. PMID:28986005
22. Huynh C, Guile J-M, Breton J-J, Desrosiers L, Cohen D, Godbout R. Is it possible to study sleep patterns in adolescent borderline personality disorder? An actigraphic feasibility study. *Int J Adolesc Med Health* 2011;22(4):547–560. [doi: 10.1515/IJAMH.2010.22.4.547]
23. Knight JF, Baber C. A tool to assess the comfort of wearable computers. *Hum Factors* 2005 Jan 1;47(1):77–91. [doi: 10.1518/0018720053653875]

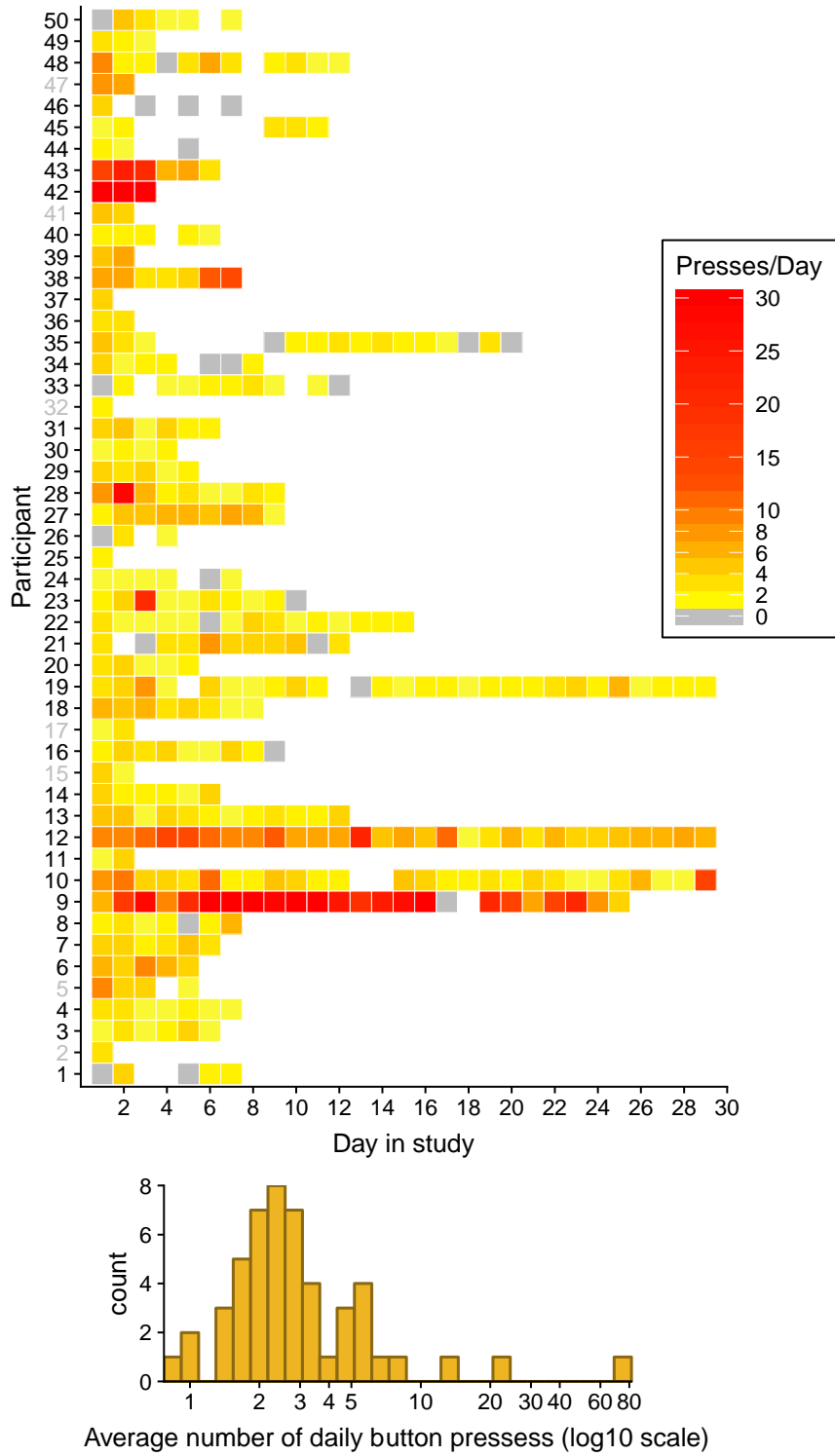
24. Picard RW, Fedor S, Ayzenberg Y. Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emot Rev* 2016 Jan 1;8(1):62–75. [doi: 10.1177/1754073914565517]
25. Empatica, Inc. What should I know to use EDA data in my experiment? [Internet]. Empatica Support. 2019 [cited 2019 Feb 15]. Available from: <http://support.empatica.com/hc/en-us/articles/203621955-What-should-I-know-to-use-EDA-data-in-my-experiment-> (Archived by WebCite® at <http://www.webcitation.org/76CtiyIb>)

Figure 1. Total daily hours worn each day (top) and histogram of average time worn each day per participant (bottom).



Notes: White squares in top panel = band not worn that day. Participants marked in grey = dropouts. For clarity, range truncated to 1-30 in top panel (7 values/1.5% of all responses > 30). Count in bottom panel refers to number of participants.

Figure 2. Total daily button presses (top) and histogram of average button presses per participant (bottom).



Notes: White squares in top panel = band not worn that day. Participants marked in grey = dropouts. For clarity, range truncated to 1-30 days in top panel (7 values/1.5% of all responses > 30).

Figure 3. Distribution of quantitative response to wearable comfort measure

