Users' Perceptions of a Digital Stress Self-monitoring Application: Research Insights to Design a Practical Innovation

Myriam Sillevis Smitt[1*], Mehdi Montakhabi[1,2], Jessica Morton[3], Cora van Leeuwen[1], Klaas Bombeke[3], An Jacobs[1]

[1]imec-SMIT, Vrije Universiteit Brussel, Pleinlaan 9, Brussels, 1050, Belgium

[2]Center of Energy Technologies (CET), Department of Business Development and Technology (BTECH), Aarhus University, Denmark

[3]imec-mict-UGent, De Krook, Miriam Makebaplein 1, Ghent, 9000, Belgium

*Correspondence: myriam.sillevissmitt@imec.be

Abstract. Self-monitoring is considered a promising tool for self-management in clinical mental health, such as for coping with excessive stress. Detecting debilitating stress before the onset of a psychopathology is becoming more of interest both for practitioners and the scientific community. However, the development of mental well-being technology focusing on stress is disrupted by the complexity of accurately measuring stress, as no clear idea exists on the construct and how it should be measured. There is also limited knowledge on the perception of perceived quality of the outcomes from a stress algorithm and the variety in its behavioural consequences. Therefore, the purpose of this study is to explore the impact of such digital self-monitoring technology for stress. It applies a qualitative method, by using semi-structured interviews. The most important resulting themes to users of this application were data-interpretation and a request for transparency. Results indicated that the majority of the predictions of the stress algorithm were not in line with the expectations of the users. The implications of these findings reveal how stress algorithms can make participants doubt their own self judgment on assessing their daily stress levels.

Keywords: Self-monitoring, Personal informatics, Mental Health, Stress, Stress algorithm

1. Introduction

It is well documented that the experience of stress, and the many forms it can manifest itself, are considered an important precursor for mental, cognitive and physical health problems [1], [2]. Stress can be triggered by negative thoughts about the past, present and future, inducing a physiological response [3]. The inability to recover from repetitive stress related events, could lead to chronic stress responses and cause diseases. The use of personal informatics, referring to personal data derived from behaviour and physiological parameters, has been considered to be a promising tool for self-management in clinical mental health, such as managing excessive stress by digital self-monitoring [4].

The detection of debilitating stress before the onset of a psychopathology has become more relevant to both practitioners and the scientific community [5], [6]. However, the development of technology to measure mental well-being is inhibited by the complexity of accurate measurement [7] as there is no clear construct of how stress, for instance, should be measured [8], [9].

Advances in wearable technology have allowed for continuous data collection. The abundance and combination of different physiological data can be used by algorithms to predict and determine behavioural patterns. The complexity of measuring stress could possibly be handled by machine learning algorithms [10]. The problem is that sensor measurements have a tendency to reduce complex phenomena to variables that can be measured, such as stress being simplified to particular heart rate patterns [8]. This is why algorithm development requires consistent and ongoing involvement with users and their data during the development of stress detection technologies.

Additionally, there is limited knowledge on how users interact with algorithmically detected stress, what the perceived quality is and what the behavioural consequences are of these stress detections [11]. Therefore, the purpose of this study is to explore the influence of digital self-monitoring for stress on participants. The core research question this study aims to answer is: 'What are users' perceptions when using a digital self-monitoring for stress tool?'

The novel contributions of this paper are 1) describing the different perceptions towards using a digital self-monitoring tool for stress, 2) demonstrating the relevance and need of interpretable algorithms for measuring stress, and 3) showing the influence of a self-monitoring for stress application on participant's their confidence in self-judgment.

The remainder of this paper is structured as follows. First, we elaborate on the current state of self-monitoring technology. Next, we introduce the relationship between self-monitoring data and algorithms. This is followed by the methodology and results of the qualitative interviews. Finally, we present a discussion and the key findings that reflect on the design and ethical implications that arose from the study.

2. Background

2.1. Self-monitoring

Self-monitoring is the collection of a person's physiological, behavioural, and physical data at repetitive intervals, in order to reveal the consistency or variation of patterns in between data points, allowing for intra or inter comparison [12]. By collecting data on oneself, self-reflection can be facilitated, providing new insights and behavioural change as a consequence [13]. Several processes are required before the data can lead to behavioural change. These processes have been developed into frameworks over the last decade and delineate the user's relationship with personal informatics and how that impacts the user [14]–[17]. All frameworks delineate a similar process, starting with creating awareness, and then reflecting and processing the data, which eventually results in an action. This reflection phase is considered to be the quintessential aspect of self-monitoring. In a qualitative systematic review on self-management for chronic diseases, a recurring theme is the “in-the-moment-understanding” that participants develop through the insights provided by self-monitored data. Seemingly, it is the skill of “in-the-moment-understanding” that users apply to associate the personal informatics with their personal beliefs of their current health [18]. Only recently a few articles shed light on the different interactions users can have with self-monitored data. There is little known about the negative consequences of (self-monitoring) data presented to a user, also known as the dark side of tracking [19]. Eikey and colleagues [11] highlight the negative thought cycles that self-monitoring can provoke, positing that limited attention has been dedicated to understanding the relationship between self-reflection in personal informatics. These negative thought cycles can lead to rumination [3], and potentially undermine the purpose of self-monitoring for health [11]. By rumination we mean recurring negative thoughts that identify with feelings of anxiety, self-doubt, insecurity or low self-efficacy. Feng and colleagues [19] conducted a systematic review on self-tracking studies and state 1) the dearth of information on how participants process the information from self-tracking and 2) the relation between self-monitoring and compulsive thoughts / behaviours.

2.2. Use of algorithms in health

Personal informatics can be used by algorithms to predict and determine behavioural patterns. Researchers and experts have shown interest in exploratory data-mining techniques to develop algorithms that may uncover new knowledge hidden within the data that cannot be seen by data inspections through the human eye.

Stress measurements are associated with several challenges that could be handled by machine learning algorithms [9]. Stress experiences are very personal, and individual and subjective differences related to stress and health perception provide another layer of potentially misleading information that algorithms can’t easily account for [20]. For example, the differences between stress measures can be related to genetic predispositions, or ethnicity-related differences [21]. Likewise, self-learning algorithms need to be trained with sensor data as well as users’ feedback to improve the accuracy of detecting stress.

However, technologies such as machine learning algorithms have come with new challenges. In particular unsupervised machine learning algorithms can have no objective for what the outcome should be, similar to measuring stress. Due to the reliance of algorithmic interpretation of data for wellbeing applications, ethical questions have been raised, such as the misleading societal implications of a non-representative sample [22]. These questions are related to every socio-technological enquiry, but their importance needs to be emphasized when used for determining human health data collected through sensors and wearables [23]. Sensors have a tendency to reduce a complex phenomena such as stress to measurable variables [8], which predominantly exposes the interpretation of sensor data and the tweaking of the algorithm feature statistics to a researchers’ bias. Although the source of data collection for a stress algorithm can be clearly stated (such as heart rate, skin conductance, temperature, accelerometer), the functionality of a stress algorithm can remain opaque to users [22]. More attention has been directed in improving the explainability and transparency of complex algorithms [24] to assess potential biases and reduce any barriers of technology adoption. The urge for more transparency and explainability has also been shown through the rise in publications on algorithms that try to win user trust.

3. Methodology

3.1 Participants

A qualitative method with semi-structured interviews was used to evaluate twelve participants' perceptions on their stress predictions. Through a purposive selection method, the interviewees were recruited from a cohort of one hundred and thirty participants taking part in the second epoch of data collection (February 2021) in a study called 'Nervocity', which investigated stress experiences in the city of Ghent. The selection of the interviewees was based on variety in education, stress levels, gender, living circumstances, and living location that represented the demographic characteristics of Ghent's citizens.

In this study, participants were excluded if they were younger than 18 years and had been diagnosed with: schizophrenia, bipolar disorder, borderline, psychotic episodes, personality disorder, or post traumatic stress disorder. In case of doubt, candidates could request a consultation with a clinical psychologist, who judged their readiness for participation. The research design was approved by the ethical commission of the research institute.

3.2 Experimental design

For two weeks, all participants wore an arm wrist wearable and answered questions about their daily stress experiences, triggered by an Ecological Momentary Assessment (EMA) methodology, on a specifically developed smartphone application for the Nervocity study (see appendix I). The data from the wearable and the EMA's were processed through imec's predictive stress algorithm [10] and visually accessible in the UI.

The wearable and wireless bracelet, called Chill+, was developed by the research company imec (see Appendix II for the technical features of the wearable). It can monitor multiple parameters through the user's wrist, such as skin conductance (measured by Galvanic skin response), heart rate (with a Photoplethysmogram, PPG sensor), temperature and movement. On the basis of these parameters and by the use of imec's algorithm [10] the degree of stress experienced by participants was estimated. These measurements were stored on the wearable and could be sent via Bluetooth for analysis, or they were read out via USB.[1]

All participants were informed on the purpose of the study and what personal data the stress algorithm used to detect stress. All study related information was written in an informed consent form and recorded in video material. No detailed information was provided on the EMA trigger methodology for questions received in the Nervocity application. Participants were informed that those smartphone questions could be linked to the wearable data, but also could be sent arbitrarily.

Interviews took place in a digital environment, through Microsoft Teams software. All interviews were recorded and during the interview the researcher took notes. After every interview the researcher wrote a summary and categorised the data by the topics discussed in the interviews (i.e., motivation, self-awareness, personal stress experience, covid impact on life, stress reduction, self-monitoring, ecological momentary assessment, user interface

[1] The Nervocity application had some boundary conditions due to the technical restrictions**.** The stress outcomes had a 15 min delay in being shown in the smartphone application, which did not allow for a realtime association between the participant's experience and the algorithm predictions. Additionally, the participants were not able to associate stress outcomes from previous days through a timeseries. This was a conscious choice by the design team as stress classification happened per minute and timeline graphs would become too variable to be interpreted at all. The technology also encompassed EMA's that were triggered through wearable data. For the purpose of this study we only address the qualitative data related to interpretation of the stress algorithm and not the EMA interactions.

experience, and use of wearable). After transcribing the topics of interest for this study, the data was coded in MAXQDA through a thematic analysis [25] and resulted in themes that reflected participants perceptions on the experiences of the stress predictions.

4. Results

4.1 Demographics

In total, 12 participants (8 female, 4 male) were interviewed, with an average age of 53 years old (SD = 12). Several relevant variables were inquired at the intake moment before the start of the study (see Table 1). Tech proficiency was very high in the current sample, measured with an average score of 4 on a 5-point Likert scale (ranging from 'totally disagree' to 'totally agree'), originating from the annual imec.digimeter report about the possession, use and attitude towards media and technology in Flanders [26]. Next, the overall happiness level reported by the current sample is moderate to high, as measured with the question "All things considered, how happy would you say you are?" on a slider scale ranging from 0 (very unhappy) to 10 (very happy). Individual stress levels were measured with the Perceived Stress Scale (PSS) inquiring experienced stress feelings and thoughts (i.e., the degree to which they feel their life has been unpredictable, out of control and overloaded) during the last month on a 5-point Likert scale (ranging from 'never' to 'always', scored 0-4) [27]. The total sum score of the 10 items answered by the current sample is considered as moderate individual stress (cut-off range between 14-26). Additionally, the average daily stress level, subjectively reported during the 2 week field study on a slider scale ranging from 1 (not stressed at all) to 10 (totally stressed), was rather low to moderate. The coping style for stress feelings during the past 2 weeks before study participation was also inquired with the Emotion Regulation Scale [28]. For each subtype of emotion regulation, three items were rated on a 5-point Likert scale (ranging from 'totally disagree' to 'totally agree', scored 1-5). Participants reported rather high integrative emotion regulation (e.g., " I mostly tried to understand why I was experiencing stress"), moderate to high suppression (e.g., "I almost always tried not to show my stress") and somewhat lower scores for dysregulation (e.g., "I sometimes did things that I really didn't want to do because of the stress"). Finally, user experience was inquired after study participation with the question "Did the study meet your expectations?" and a 5-point Likert scale, resulting in rather mixed answers and thus a neutral average. Also, the User Burden Scale was used to get an idea of the user burden when using the Nervocity technology (wearable + application) with 20 items on a 5-point Likert scale (ranging from 'never' to 'always', scored 0-4). These items touched upon the physical, time and social, mental and emotional and privacy burden, and general difficulty of use [29]. In general, experienced user burden was very low.

Table 1. Summary of demographic and context variables measured before, during and after study participation.

Variable	**Mean (SD) / N (%)**
Gender	8 Females (67%)
Age	53 years (12)
Tech proficiency (1:5)	4.29 (0.69)
Happiness level (0:10)	6.92 (1.0)

Perceived stress (0:40)	17.33 (5.37)
Daily subjective stress level (1:10)	3.78 (1.67)
Integrative emotion regulation (1:5)	3.75 (0.62)
Suppressive emotion regulation (1:5)	2.92 (0.79)
Dysregulation (1:5)	2.47 (0.90)
Study expectations met (1:5)	3.33 (0.89)
User burden (0:80)	13.58 (7.57)

The thematic analysis yielded two main themes: “data interpretation” and “request for transparency”. These themes were further divided in five subthemes (figure 1 shows a brief overview of the theme structure). The sub-theme “Discrepancy between self evaluation and algorithm” is the most substantial of the (sub)themes and further consists of two distinct categories “the ruminators” and “the rejectors”, explained in more detail below. These are preliminary findings and are subject to change and recategorization when the other additional thirteen interviews from the Nervocity study have been analysed.

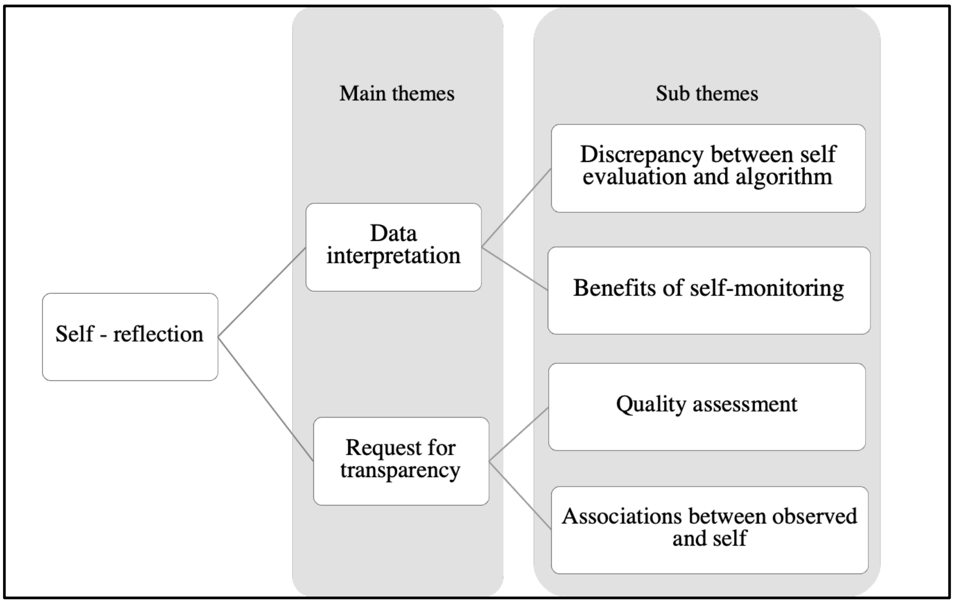

Figure 1. Qualitative data theme overview

4.2 Data-interpretation theme

Discrepancy between self evaluation of stress and algorithmic stress. The majority of the participants experienced a discrepancy between their perception of daily experienced stress and the stress predictions coming from the stress algorithm. The most frequent occurring comment was that the algorithm categorized stress high(er) and more frequent in contrast to the participant’s perception: “It was indicated that I was stressed while I was actually feeling very good.” (P01) and “I don't like it when the wearable says something different from what I experience.” (P02). Due to this perceived disparity it was possible to discern two types of participants, being 1) ruminator and 2) rejector of the technology.

These types were based on the behaviour and reported reactions towards the algorithmic detected stress.

As the name implies, the "ruminators" are the group of participants that experienced a lot of worry associated with the predicted stress. They tended to accept the indicated stress level and this resulted in them questioning their own perception and judgement. Furthermore, this resulted in feelings of insecurity, as they were mostly concerned about what they should do with this information: "If I showed the app to my friends, they would tell me that I should look for another job." (P03).

Indeed, these participants would overanalyse the results and tended to trust the findings of the algorithm more than their own perception of experienced stress levels: "The app indicated much more stress than I felt. Is it the app or is it my perception?" (P02). Another example is: "I started to think that maybe it was just me. I know that I am sensitive to tension and stress, but participation still makes me doubt, should I do something with this information now?" (P03). This confusion was experienced by ruminators in regard to their supposed actions towards the detected stress: "The wearable registers stress at times when I think I don't experience stress. I do not know to what extent it is correct." (P04) and "I have stress [referring to technology] when I don't expect it, and I have no solution for it." (P05).

The cognitive dissonance between the self perceived and the algorithmic stress outcome seemed to be related to the differences in initial motivation to join the study. For instance, participants expressed that they had certain expectations, such as confirming their self-perception of low or high stress by use of technology: "The theme interests me. I want to gain more insight into my own stress. I want to know if my feelings and experiences match the wearable." (P03) and "I know myself that I regularly experience stress. I've been thinking about the idea of buying a smartwatch for a while now. By joining the study I thought it would be a good idea to see if it works effectively and what I think of it." (P04). Another example is "In the beginning I did not trust the system, but because the second week did not differ from the first week and the type and frequency of the recorded stress was similar, I started to think that It was me after all." (P03). The participants expressed being sensitive to experiencing stress and partaking in this study resulted in more doubts: "But do I do something with this information [stress detection]?" (P03).

The second group of participants decided to reject the technology as the discrepancy between their own stress experience and the algorithmic findings resulted in a disbelief of distrust in the technology. This group was therefore called "the rejectors of technology". The rejection was a result of a fair bit of reflection, for instance if the participant started to worry at first due to high recorded stress, however instead of adopting the results from the algorithm she decided to take distance from the self-monitored data: "I thought it was very frustrating, and started to worry. I started to wonder if I am so stressed that I can't even detect it myself anymore." (P08). After a week she came to the conclusion that it was not her fault but the system. Then she was able to let it go emotionally.

For another participant the initial motivation was to prove herself how successful her stress reduction techniques were. However, when she experienced that the stress data was not in line with her perception she got annoyed: "I only see orange [high stress] while I think I do not experience any stress." (P07). Later in the conversation, she indicated that this made her a bit angry: "I feel a bit resented." (P07). Eventually, she dismissed the data completely: "I am not experiencing any stress at the moment and I think the study did not measure any stress either." (P07).

Rejecting the personal informatics can take place due to disagreement, but can also be a means to avoid thoughts of rumination and self-doubts. One participant mentioned she was avoiding looking at the data. She describes that when she is not experiencing stress but the stress algorithm is, she does not want to know: “If it doesn’t bother me, it is better to leave it that way.” (P11). Implying that she rather did not want to look at the user interface anymore. She even expressed that the difference between her own opinion and the algorithm outcome was quite funny: “Sometimes I was just busy with work, household work, then it indicated that as a stress peak, while I didn't think it was. I thought it was comedic." (P11).

In contrast to the discrepancy between self-perception and the algorithmic outcomes, one participant did not question the self-monitored data and identified completely with the stress outcomes: “It confirmed to me that I am quickly triggered by stress.” (P09).

Benefits of self-monitoring. The use of self-monitoring for stress made several participants reflect on their behaviour, emotions and cognition. Although several participants expressed doubts about the accuracy of the algorithmic predictions, partaking in the study did have advantages that were independent from the evaluation of the technology. Regularly, self-monitoring studies have delineated the influence of personal informatics. In this study we found that participants expressed more 1) self-awareness, such as “I have been listening more to my body” (P04), 2) better recognition of behavioural patterns when stressed, such as “I became more aware of the hours in the day and what I did at that time because of the study.” (P04), referring to the smartphone questions about their stress experience, and 3) a better sense of time and activities due to the reflective questions. It made participants reflect on their day and on their emotions more than they would do without the use of the wearable, together with the triggered smartphone questions: “If I think about it, I act differently when I am stressed compared to when I am at ease, I am more on edge when I am stressed, otherwise I never have that.” (P09).

4.3 Request for transparency theme

Associations between observed and self. Regardless of the type of emotional response the participants had to their stress predictions, most of them expressed a curiosity to understand how stress can be determined. For instance, participants expressed an interest in wanting to know how their heart rate was associated with stress: “I want to be able to make associations between the data and my own experiences.” (P06). Or participants looked for explanations themselves: “I think that when my wrist is on my chest when I am resting, it raises the temperature and measures high stress. I think that's not right. Similarly, there were times when the sun was shining on the watch, causing the temperature to rise and incorrectly measuring stress and making me sweat a lot. So every time I am active, stress is measured.” (P12). Another participant interpreted the stress predictions as positive stress: “I came to the conclusion that I did not understand that stress was detected while I did not experience it and so I thought it was interpreted as positive stress, but I do not think that was correct.” (P01).

Quality assessment of algorithm. Participants who expressed a discrepancy between the self perceived stress and the algorithm outcome, wondered about the quality of the stress algorithm. The stress predictions were considered to be of interest, but due to the distrust in the stress algorithm, participants felt the need to identify the reliability of the algorithm. There was a need for accountability and explainability of the algorithmic predictions. As one participant stated he finds it difficult to deal with the high stress values. If he understands better what the cause of it is or how it is detected he would feel more at ease: “I want to know more about how stress is determined, what is the cause? If I can understand it better, then I know how to deal with it better.” (P05).

Another participant stated that she thought the frequency of her high stress prediction was very excessive: “I am in the orange zone [high stress] a lot, I wonder if this is correct. I thought to myself more to be in the yellow zone [light stress]. I started to question what has actually been measured?” (P04). She questioned the possibility of measuring stress: “Only two parameters are measured, temperature and heart rate, how can that measure stress? Is that possible?” (P04).

Also the question for comparing own predictions with that of other participants was raised: “I would like to be able to compare my data with that of others, do they also register that much stress? I also would like to know more about how they detect stress, so I can better assess if it actually is stress or not. Now I have no clue and it makes me doubt. Although I know I am sensitive to stress…but is it really the case? I can’t interpret it very well.” (P03).

A few participants did not articulate strong thoughts on the algorithmic stress outcomes. It seemed that the lack of interpretability of the data resulted in assuming that the technology is still in development: “I assumed that it is a study and that it [stress algorithm] is still under development.” (P12). One participant expressed that human interpretation is not always correct and sometimes technology is flawed: “It is an interplay, it depends from situation to situation. Sometimes the wearable is wrong, sometimes I'm wrong. I assume that technology is not always correct and that we are not robots.” (P06).

5. Discussion

This study has shown that the algorithm categorised higher stress than participants perceived to be correct, leading to differing attitudes. The primary attitudes that were observed were participants who kept ruminating, those who rejected algorithmic stress detection and a few who remained indifferent. It is this variety of the different attitudes towards the use of the stress algorithm that provided new insights and further discussion for its future development.

It is only recently that studies demonstrated the dark side of self monitoring [11], [19]. They have shown that too much emphasis on individuals' thoughts can impede their mental health. Eikey and colleagues [11] pointed out that the difference between negative thought cycles (such as rumination) and constructive self-reflection could relate to parameters such as motivation, goals that identify with the user's self-worth and culturally valued morals. Stress continues to have a negative connotation and the purpose in western societies remains to reduce stress or to improve oneself to better deal with stress [30]. Participants focusing on determining how high their stress levels are can potentially demonstrate the emotional battle they have with the concept of stress and their work life balance.

In the current data, we found that several participants wanted to confirm their own stress observations, which could indicate that they initially doubted or questioned their own judgment of stress perception. Participants showed a high interest in their own stress levels, reflected by high integrative emotion regulation scores (see Table 1). Commencing the study with high levels of self-doubts or low self-efficacy on their stress judgment could be of influence in how participants interpreted the stress detections from the algorithm. Although for some participants the algorithmic stress detection confirmed their initial ideas, the frequency of high stress that was recorded for them could have felt like a moral judgment and made them question themselves or the technology.

Lupton [31] and Feng and colleagues [19] mention the neoliberal shift to feeling responsible for one’s individual health and well-being. Commercial digital health innovations are built on the premise of improving well-being, and tracking technologies imply the individual as deficient in being knowledgeable enough to make health and well-being related decisions [23]. This implicit tendency could possibly influence the relationship users nowadays have with technology and makes them susceptible to depend on algorithmic data to confirm the beliefs one has about oneself. Interestingly enough, a few participants pointed out the experiences of self-doubt when being confronted with the stress predictions, yet decided to be in disagreement with this algorithmic stress output at some point.

The secondary theme in this study originated from the discrepancy participants experienced with their stress predictions. All participants were informed of the type of data collection to determine the stress outcome and were aware that the answers to the smartphone questions, not further discussed in this article, helped them to further personalise their stress predictions. In order to detect biases or to determine the trustworthiness in the algorithmic stress predictions, participants followed their own explanations or expressed their need for explainability of the algorithm. Verifying and understanding an algorithm for personal use is an important aspect for adopting the technology [32].

Independent from how participants interpreted their detected stress, many expressed that reflecting on their stress benefited their own self-awareness. This is in line with many other studies that use self-monitoring for well-being [13], [33]–[35].

5.1 The insights to design a self-monitoring practical innovation

Technology developers should be aware of the behavioural implications of applications that show personal informatics and algorithmic predictions. Although the characteristics of the group of participants that experienced rumination is unclear, it is suggested to take the potential negative consequences of personal informatics into account in the design of mental well-being technologies. Additional emphasis must be placed on the differences in data interpretation and if technology meets the initial purpose of its design. A few aspects need to be considered when designing user interfaces that show well-being related algorithmic predictions:

- A need for greater explainability when using algorithmic predictions related to health. This will help to increase the users confidence in deciding to trust or distrust the algorithm.
- Increase interpretability through developing a comparative framework, such as linking the algorithmic predictions of users to their notes and their calendar activities.
- Increase interpretability and trustworthiness by developing normality curves, enabling anonymous data comparison between other users.
- Defining 1) the user's motivation to use the technology, 2) their self-efficacy in self-evaluation, and 3) their perceived stress levels in order to shape the type of information and communication style that is provided in the design of the UI.

5.2 Limitations and Opportunities for Future Research

The current study has limitations and therefore opportunities for further research. The study had a small sample size and was performed during Covid restrictions. This could be remedied in the future by a larger sample size which represents society more broadly and in more representative living circumstances.

Measuring stress is subject to individual interpretation. For this study we provided a broad description of stress as the debilitating effect of stress, which is very subjective. The qualitative data, not discussed in the findings of this study, demonstrated a homogeneous interpretation that stress is predominantly a negative term [20]. However, when participants felt unable to associate the predicted algorithmic stress data with their own perception, they classified the predicted stress as a form of positive stress or as physiological stress through exercising. Therefore, studying the ontology of stress and its associations with psychological traits is required for future research.

Many interviewed participants who participated in the Nervocity study expressed an interest in technology or expressed a curiosity in understanding their stress levels. Due to this selection bias, it could be that participants were more prone to be influenced by the outcomes of the stress algorithm. Future research should be aware of the impact of stress predicting algorithms and account for potential adverse effects. A shift in focus is also needed to determine what kind of psychosocial parameters are related to having a debilitating interaction instead of a constructive relationship with self-monitoring for stress.

The UI design of the smartphone application undoubtedly had an influence on how participants perceived and interacted with the stress detections. The technical limitations did not allow for more granularity and variation in design choices. Aspects that were of influence were: colour choice, wording, timeline format, etc. Future research should perform A/B testing on different types of UI designs while involving end users.

Participant involvement required two weeks of usage of the technology and interacting with stress predictions. This short time window can not establish stable behavioural patterns when it comes to the usage and interpretation of predicted stress outcomes. The novelty and motivation overruled the honest judgement of the technology. Longitudinal data would demonstrate differences more thoroughly when it comes to evaluation of the stress data and how it could have impacted their lives.

Finally, although the technology for the stress algorithm is a self-learning algorithm, the initial dataset for this algorithm was developed on a white-collar dominated data set and therefore biases could be present. For instance, one of the most recurring observations by the participants was high stress during office hours or during long periods of sitting. Participants consequently associated their predicted high stress with work activities and assumed that their work caused more stress than they were able to realise themselves. This could be true, but could be a default feature of the previous trained dataset of this algorithm, with false assumptions as a consequence. Qualitative user input is required to highlight contextual situations that could improve and correct the efficacy of the predictive model.

6. Key insights

- Participants had a tendency to consider the predicted stress outcomes being more objective than their own self-evaluation of stress. This resulted in self-doubt or rejection of the technology. The important take-away message here is the influence technology can have, regardless of its reliability.
- Using a stress prediction algorithm can help to reflect on daily behaviour patterns and create more self-awareness. This is independent from the perceived reliability of the algorithm.
- The type of intention or motivation for using a stress algorithm could possibly be associated with negative thought cycles and lead to rumination and aggravating mental well-being.
- Participants that expressed a low interpretability of the predicted stress measures, resulted in a distrust in the algorithm or development of their own theories to explain the data. More explainability of the algorithm is required to support the user's decision making if and when an algorithmic stress prediction can be trusted and therefore utilised.

7. Concluding Remarks

The study explored users' interaction with digital self-monitoring for stress. Based on the results of twelve semi-structured interviews, two main themes were derived: 'data interpretation' and 'request for transparency'. A majority of the predictions of the stress algorithm were not in line with the expectations of the participants. This resulted in a high level of self-reflection that instead of providing insight, created more self-doubt or adversity towards the technology. Participants seemed to rely on the algorithmic stress detections and started questioning the capability of their own judgment when assessing their own mental health.

Acknowledgements

This work was supported in part by VLAIO, Stad GENT, imec through the Nervocity project. Mehdi Montakhabi is funded by the Flemish Government through the FWO-SBO project SNIPPET-S007619. Cora van Leeuwen is funded by SBO-grant S005221N, funded by The Research Foundation – Flanders. The authors would like to thank Tim Coggins and Shirley Elprama for proofreading.

References

[1] M.-F. Marin *et al.*, 'Chronic stress, cognitive functioning and mental health', *Neurobiol. Learn. Mem.*, vol. 96, no. 4, pp. 583–595, Nov. 2011, doi: 10.1016/j.nlm.2011.02.016.

[2] S. Richardson, J. A. Shaffer, L. Falzon, D. Krupka, K. W. Davidson, and D. Edmondson, 'Meta-Analysis of Perceived Stress and Its Association With Incident Coronary Heart Disease', *Am. J. Cardiol.*, vol. 110, no. 12, pp. 1711–1716, Dec. 2012, doi: 10.1016/j.amjcard.2012.08.004.

[3] J. F. Brosschot, 'Markers of chronic stress: Prolonged physiological activation and (un)conscious perseverative cognition', *Neurosci. Biobehav. Rev.*, vol. 35, no. 1, pp. 46–50, Sep. 2010, doi: 10.1016/j.neubiorev.2010.01.004.

[4] E. L. Murnane *et al.*, 'Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health', *J. Am. Med. Inform. Assoc.*, vol. 23, no. 3, pp. 477–484, May 2016, doi: 10.1093/jamia/ocv165.

[5] G. H. Crossley, A. Boyle, H. Vitense, Y. Chang, and R. H. Mead, 'The CONNECT (Clinical Evaluation of Remote Notification to Reduce Time to Clinical Decision) Trial', *J. Am. Coll. Cardiol.*, vol. 57, no. 10, pp. 1181–1189, Mar. 2011, doi: 10.1016/j.jacc.2010.12.012.

[6] J. Firth, J. Torous, and A. R. Yung, 'Ecological momentary assessment and beyond: The rising interest in e-mental health research', *J. Psychiatr. Res.*, vol. 80, pp. 3–4, Sep. 2016, doi: 10.1016/j.jpsychires.2016.05.002.

[7] M. Reisinger and K. Röderer, 'I'm fine, thank you – Contextualizing Wellbeing and Mental Health for Persuasive Technologies', p. 6.

[8] J. Müller, S. Fàbregues, E. A. Guenther, and M. J. Romano, 'Using Sensors in Organizational Research—Clarifying Rationales and Validation Challenges for Mixed Methods', *Front. Psychol.*, vol. 10, p. 1188, May 2019, doi: 10.3389/fpsyg.2019.01188.

[9] E. Smets, W. De Raedt, and C. Van Hoof, 'Into the Wild: The Challenges of Physiological Stress Detection in Laboratory and Ambulatory Settings', *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 463–473, Mar. 2019, doi: 10.1109/JBHI.2018.2883751.

[10] E. Smets, 'Towards large-scale physiological stress detection in an ambulant environment', p. 198.

[11] E. V. Eikey *et al.*, 'Beyond self-reflection: introducing the concept of rumination in personal informatics', *Pers. Ubiquitous Comput.*, vol. 25, no. 3, pp. 601–616, Jun. 2021, doi: 10.1007/s00779-021-01573-w.

[12] A. J. Karter *et al.*, 'Longitudinal Study of New and Prevalent Use of Self-Monitoring of Blood Glucose', *Diabetes Care*, vol. 29, no. 8, pp. 1757–1763, Aug. 2006, doi: 10.2337/dc06-2073.

[13] S. Compernolle *et al.*, 'Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis', *Int. J. Behav. Nutr. Phys. Act.*, vol. 16, no. 1, p. 63, Dec. 2019, doi: 10.1186/s12966-019-0824-3.

[14] E. P. S. Baumer, 'Reflective Informatics: Conceptual Dimensions for Designing Technologies of Reflection', in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul Republic of Korea, Apr. 2015, pp. 585–594. doi: 10.1145/2702123.2702234.

[15] G. I. Wolf and M. De Groot, 'A Conceptual Framework for Personal Science', *Front.*

Comput. Sci., vol. 2, p. 21, Jun. 2020, doi: 10.3389/fcomp.2020.00021.
[16] I. Li, A. Dey, and J. Forlizzi, 'A stage-based model of personal informatics systems', p. 10, 2010.
[17] D. A. Epstein, A. Ping, J. Fogarty, and S. A. Munson, 'A lived informatics model of personal informatics', in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, Osaka, Japan, 2015, pp. 731–742. doi: 10.1145/2750858.2804250.
[18] S. Tadas and D. Coyle, 'Barriers to and Facilitators of Technology in Cardiac Rehabilitation and Self-Management: Systematic Qualitative Grounded Theory Review', *J. Med. Internet Res.*, vol. 22, no. 11, p. e18025, Nov. 2020, doi: 10.2196/18025.
[19] S. Feng, M. Mäntymäki, A. Dhir, and H. Salmela, 'How Self-tracking and the Quantified Self Promote Health and Well-being: Systematic Review', *J. Med. Internet Res.*, vol. 23, no. 9, p. e25171, Sep. 2021, doi: 10.2196/25171.
[20] S. Sharma, G. Singh, and M. Sharma, 'A comprehensive review and analysis of supervised-learning and soft computing techniques for stress diagnosis in humans', *Comput. Biol. Med.*, vol. 134, p. 104450, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104450.
[21] A. K. Manrai *et al.*, 'Genetic Misdiagnoses and the Potential for Health Disparities', *N. Engl. J. Med.*, vol. 375, no. 7, pp. 655–665, Aug. 2016, doi: 10.1056/NEJMsa1507092.
[22] F. Cabitza, R. Rasoini, and G. F. Gensini, 'Unintended Consequences of Machine Learning in Medicine', *JAMA*, vol. 318, no. 6, p. 517, Aug. 2017, doi: 10.1001/jama.2017.7797.
[23] D. Lupton, 'Health promotion in the digital era: a critical commentary', *Health Promot. Int.*, vol. 30, no. 1, pp. 174–183, Mar. 2015, doi: 10.1093/heapro/dau091.
[24] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, 'From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices', *Sci. Eng. Ethics*, Dec. 2019, doi: 10.1007/s11948-019-00165-5.
[25] V. Braun and V. Clarke, 'Using thematic analysis in psychology', *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, Jan. 2006, doi: 10.1191/1478088706qp063oa.
[26] K. Vandendriessche, E. Steenberghs, A. Matheve, A. Georges, and L. De Marez, 'imec. digimeter 2020: Digitale Trends in Vlaanderen', 2021, [Online]. Available: https://biblio.ugent.be/publication/8717212/file/8717464
[27] S. Cohen, T. Kamarck, and R. Mermelstein, 'A Global Measure of Perceived Stress', *J. Health Soc. Behav.*, vol. 24, no. 4, p. 385, Dec. 1983, doi: 10.2307/2136404.
[28] G. Roth, A. Assor, C. P. Niemiec, R. M. Ryan, and E. L. Deci, 'The emotional and academic consequences of parental conditional regard: Comparing conditional positive regard, conditional negative regard, and autonomy support as parenting practices.', *Dev. Psychol.*, vol. 45, no. 4, pp. 1119–1142, 2009, doi: 10.1037/a0015272.
[29] H. Suh, N. Shahriaree, E. B. Hekler, and J. A. Kientz, 'Developing and Validating the User Burden Scale: A Tool for Assessing User Burden in Computing Systems', in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, Santa Clara, California, USA, 2016, pp. 3988–3999. doi: 10.1145/2858036.2858448.
[30] T. T. Donnelly and B. C. Long, 'Stress Discourse and Western Biomedical Ideology: Rewriting Stress', *Issues Ment. Health Nurs.*, vol. 24, no. 4, pp. 397–408, Jan. 2003, doi: 10.1080/01612840305316.
[31] D. Lupton, 'The digitally engaged patient: Self-monitoring and self-care in the digital health era', *Soc. Theory Health*, vol. 11, no. 3, pp. 256–270, Aug. 2013, doi: 10.1057/sth.2013.10.
[32] V. Alexander, C. Blinder, and P. J. Zak, 'Why trust an algorithm? Performance, cognition, and neurophysiology', *Comput. Hum. Behav.*, vol. 89, pp. 279–288, Dec. 2018, doi: 10.1016/j.chb.2018.07.026.
[33] E. P. Bray, R. Holder, J. Mant, and R. J. McManus, 'Does self-monitoring reduce blood pressure? Meta-analysis with meta-regression of randomized controlled trials', *Ann. Med.*, vol. 42, no. 5, pp. 371–386, Jul. 2010, doi: 10.3109/07853890.2010.489567.
[34] B. R. Fletcher, J. Hartmann-Boyce, L. Hinton, and R. J. McManus, 'The Effect of Self-Monitoring of Blood Pressure on Medication Adherence and Lifestyle Factors: A Systematic Review and Meta-Analysis', *Am. J. Hypertens.*, vol. 28, no. 10, pp. 1209–1221, Oct. 2015, doi: 10.1093/ajh/hpv008.
[35] Y. Kanejima, M. Kitamura, and K. P. Izawa, 'Self-monitoring to increase physical activity in patients with cardiovascular disease: a systematic review and meta-analysis', *Aging Clin. Exp. Res.*, vol. 31, no. 2, pp. 163–173, Feb. 2019, doi: 10.1007/s40520-018-0960-7.

APPENDIX I

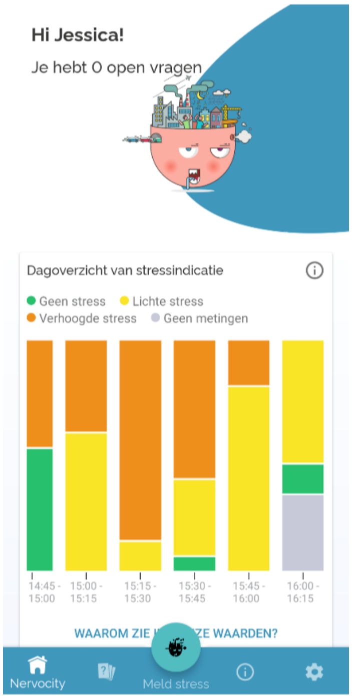

Figure 2. User interface visualisations: visual on the left show stress prediction during the day, visual on the right shows aggregated stress during the week.

APPENDIX II

Technical specifications wearable Chill+

	Parameter	Details	Units	
Physiological Signals	GSR	Dynamic Range	1 - 20	µS
		Dynamic Range	0.01 – 2	µS
		Sampling frequency	Configurable (16-256)	Hz
		Resolution	16	bits
	PPG	Dynamic Range	50	µA
		Resolution	22	bits
		Sampling frequency	Configurable (64 -256)	Hz
	Skin temperature	Dynamic Range	10 - 45	^{0}C
		Sampling frequency	1	Hz
		Resolution	12	bits
Contextual Signals	Accelerometer	Dynamic Range	±2	g
		Sampling frequency	32	Hz
		Resolution	16	bits
	Gyroscope	Dynamic Range	±250	(°/sec)
		Sampling Frequency	32	Hz
		Resolution	16	bits

	Details	Units	
System Parameters	Operating voltage	3 - 4.2	V
	Charging time	90	minutes
	Autonomy (Storage)	24	hours
	Operating Mode	Storage and Streaming	-