

# Unpacking Health Data: How Metrics Shape Personal Informatics Experiences

Marit Bentvelzen  
Utrecht University  
Utrecht, The Netherlands

## Abstract

Personal informatics systems, such as wearables and smartphone applications, help individuals track data to reflect on their health and well-being. Personal informatics models describe how users interact with their health data at a meta-level. Our recent work highlights that this perspective lacks sufficient granularity, as users' perceptions, understanding, and emotional responses vary across different metrics. In this position paper, I synthesise findings from two recent studies to introduce four key dimensions of health metrics: *abstraction*, *sensitivity*, *transparency*, and *actionability*. I discuss how these dimensions shape users' experiences with personal informatics and outline their implications for designing future health-tracking systems. By unpacking these dimensions, we can better align system design with users' diverse needs, fostering more meaningful engagement with health data.

## Keywords

Metrics, Health data, Health & Well-being, Self-tracking, Personal Informatics

## ACM Reference Format:

Marit Bentvelzen. 2025. Unpacking Health Data: How Metrics Shape Personal Informatics Experiences. In *Proceedings of (CHI '25 Workshop on Envisioning the Future of Interactive Health)*. ACM, New York, NY, USA, 3 pages.

## 1 Introduction

People often turn to technology to keep track of their health and well-being. From fitness trackers, to smart scales, blood pressure monitors and smartphone apps. Such devices are often referred to as Personal Informatics (PI) [6] which are systems that help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge. In PI, the users' journey and lived experience are often described through models [4]. Currently, these models describe how users interact with their (health) data from a meta perspective, considering personal data as a black box.

Health is tracked using a wide variety of metrics —steps, sleep quality, stress, body battery, blood pressure, VO2max to name a few. Our recent work [1, 10] suggests that how users perceive, feel about, and understand their data differs per metric. As such, it is necessary to further unpack health data and explore metrics in more detail.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI '25 Workshop on Envisioning the Future of Interactive Health, Yokohama, Japan  
© 2025 Copyright held by the owner/author(s).

Drawing on two recent studies I explore four dimensions of metrics: *abstraction*, *sensitivity*, *transparency*, and *actionability*, and discuss how these dimensions shape users' experiences with their health data. Subsequently, I discuss the implication of these dimensions for the design of health & well-being technology, demonstrating that design patterns for self-reflection [2], design principles in persuasive technologies [8], and the game elements in gamification [5], may not be suitable for every metric.

This position paper offers a starting point for discussing the different dimensions of metrics. By unpacking these dimensions, I hope to inspire future research on how health metrics can be designed and applied more effectively to support users' health & well-being.

## 2 Background: Studies on Metrics

In recent years, we conducted several studies to learn how people perceive different health-related metrics. In the first [1], we looked at how users perceive (derived) metrics. We conducted an online vignette survey study ( $n = 228$ ) and an interview study ( $n = 12$ ) comparing a fictional health metric presented at three different levels of abstraction. Participants found it hard to understand a generalised *health score*, questioning how it was calculated, what it measured precisely, and how to interpret a percentage-based score. They only started to engage with the health score, trying to interpret it, when additional metrics (e.g., resting heart rate, stress level, steps) were shown alongside it. Participants said they wanted more transparency about how the metric was defined and calculated and more guidance on improving their scores. These concerns mainly arose for more abstract metrics like *health*, *stress*, or *body battery*. In contrast, participants did not question the meaning or measurement of metrics such as step count or resting heart rate, suggesting that interpretability differs depending on the type of metric.

In our most recent study [10], we examined how social comparisons (e.g., leaderboards) affect users' affective responses to health data. First, we reviewed the Top 50 Health & Wellness smartphone apps from data.ai<sup>1</sup> to identify the social comparison strategies they use and the metrics involved. Of the 146 distinct metrics found, 71 were used for social comparisons, ranging from *steps* or *total minutes active* to more sensitive metrics like *BMI*, *body fat percentage*, or *calories consumed*. We then conducted an online vignette study ( $n = 192$ ) and an interview study ( $n = 12$ ). For the vignette study, we designed several interface prototypes that used different comparison strategies (e.g., showing the user's data on a distribution, in a feed, or on a leaderboard), using *steps* and *body fat percentage* as example metrics. Using the Positive and Negative Affect Schedule (PANAS), we found that social comparisons can motivate users

<sup>1</sup>See supplementary materials of our paper [10] for the full list of apps

and elicit positive emotions (e.g., pride, enthusiasm) but can also lead to negative feelings (e.g., shame, guilt). The type of metric and comparison strategy both mattered: *steps* tended to increase positive affect, while *body fat percentage* prompted more negative affect—particularly when users were ranked low on a leaderboard.

These findings show that some metrics are more challenging to interpret or more sensitive than others, and people have different expectations about how certain metrics should be explained or contextualised. This highlights the need to look at personal health data in more detail and to consider how differences between metrics affect people's understanding, affective response, and interpretation. Based on these findings, I have identified four dimensions that may explain the differences in how users perceive metrics: abstraction, sensitivity, transparency, and actionability.

### 3 Dimensions of Health & Well-being Metrics

#### 3.1 Abstraction of Metrics

The first dimension is the *abstraction* of metrics. We previously explored this dimension, building on Construal Level Theory (CLT) from Trope et al. [9]. This theory differentiates between high level (i.e., relatively abstract), and low level (i.e., relatively specific) construal; and indicates that this abstraction level influences the cognitive processing of information. The theory notes that physical activities can be construed either in a more abstract (e.g. focusing on the bigger picture) or more concrete manner (e.g. focusing on how an activity is performed). Hence, CLT could offer a lens through which we can view the abstraction levels of metrics. In this study [1] we assumed *health* to be a more abstract metric than *steps* and *heart rate*, and our findings indeed indicate that participants needed additional metrics to make sense of a *health* score. However, at this point we do not fully grasp what makes a metric more or less abstract, and later on, when further reflecting on this study, I realised that there may be different attributes to a metric that could make the metric more or less abstract: (1) The meaning of a metric might be abstract, or (2) How the metric relates to health, or (3) How it is measured, (4) The unit of measure may be more abstract—for instance, *VO2max* is expressed in milliliters of oxygen per kilogram of the body mass per minute (ml/kg/min), or (5) A metric might be less abstract when you can sense it yourself (this may explain why our participants did not question how steps were detected by their fitness tracker).

#### 3.2 Sensitivity of Metrics

Our study on social comparisons [10] shows that health metrics are not objective or neutral entities. Certain metrics seem to be perceived as more *sensitive* than others, and may evoke more shame or touch on a taboo. Previous work by Murnane et al. [7] also touched on this, their ecological model illustrates how interactions within a user's wider social context shapes their experiences with personal informatics systems, sometimes reinforcing negative emotions such as shame and rejection due to societal stigmas, institutional practices, and cultural expectations. Our work [10] shows that participants considered *body fat percentage* to be much more sensitive than *steps*, yet we only explored these two metrics, and do not know which metrics are perceived to be more sensitive and

why. To address this gap, we are currently exploring the sensitivity of all metrics found in the Top 50 of Health & Wellness apps.

#### 3.3 Transparency of Metrics

The perceived *transparency* regarding the definition or measurement of a metric forms another dimension. In general, wearable companies such as Fitbit, Garmin, and Whoop describe metrics on a high level and do not provide specific definitions for metrics or transparency regarding the algorithms underlying the metric. However, as mentioned in section 2, people do not require transparency for each metric. Based on our studies, I hypothesise that transparency becomes more relevant when a metric is multi-interpretable. For instance, a high-level concept such as health could be interpreted and operationalised in many ways. Similarly, a *stress* metric could be interpreted as physiological or psychological stress. Previous work has shown that users of fitness trackers had a hard time making sense of their trackers' stress scores, as this reflected physiological stress. In contrast, users conceptualise stress as a psychological phenomenon [3]. As such, the need for transparency is also related to the abstraction of a metric.

#### 3.4 Actionability of Metrics

Finally, *actionability* forms a dimension of metrics. In our study on derived metrics [1], participants considered actionability to be more valuable than transparency. Understanding how a metric is defined and calculated does not offer insight into what and how to improve. There are metrics for which actionability seems more straightforward: to improve the *number of steps* a person needs to walk or run more; to improve *water intake*, a person should drink more water; and to improve *total active time*, a person should do more physical activity. Yet, for other metrics, actionability is less straightforward; for instance, participants often indicated that they had no idea how to improve their *VO2max*, *body battery*, and *sleep quality*. There are specific medical guidelines for some metrics, such as *blood pressure*, yet for many metrics, such guidelines do not exist.

Although these four dimensions provide a foundation for discussing and exploring differences between metrics, they are by no means exhaustive. I hope they will inspire a more detailed exploration of health & well-being metrics.

### 4 What Do These Dimensions Imply for the Design of Health Informatics Systems?

These dimensions suggest we must engage with data at the level of individual metrics, rather than treating personal data as a catch-all. In our previous work [2], we introduced 11 strategies and 74 design patterns to enhance reflection in personal informatics and guide reflective experience design. However, we focused broadly on personal data, overlooking the nuances of individual metrics. For instance, using competitive comparisons (i.e., leaderboards) for a *sensitive* metric (e.g., body fat percentage) can provoke shame and distress, whereas using a design resource such as *ambiguity* will probably not be beneficial for an abstract metric. Asking a user reflective questions (i.e., *What actions could you take to improve your body battery this week?*) may not be useful for metrics where

*actionability* is less obvious or straightforward. Consequently, design patterns for self-reflection [2], design principles in persuasive technologies [8], and the game elements in gamification [5], may not be suitable for every metric. Thus, we need to further unpack health data and explore the different dimensions of metrics to design health informatics systems that fit users needs.

## 5 Conclusion & Future Work

In this position paper, I explore how health-related metrics—such as steps, heart rate, or body fat percentage—differ across dimensions like abstraction, sensitivity, transparency, and actionability. Recognising these nuances allows us to move beyond a one-size-fits-all approach and create more personalised, meaningful health informatics systems. For future work, I will investigate which metrics people find most sensitive or complex to act upon, and why, so that we can design interfaces tailored to each metric’s specific needs. By refining how personal health data is presented, we can offer users more meaningful support for tracking their health & well-being.

## References

- [1] Marit Bentvelzen, Jasmin Niess, and Paweł W. Woźniak. 2023. Designing Reflective Derived Metrics for Fitness Trackers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 158 (jan 2023), 19 pages. doi:10.1145/3569475
- [2] Marit Bentvelzen, Paweł W. Woźniak, Pia S.F. Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting Reflection in HCI: Four Design Resources for Technologies that Support Reflection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (March 2022), 2:1–2:27. doi:10.1145/3517233
- [3] Xianghua (Sharon) Ding, Shuhan Wei, Xinning Gui, Ning Gu, and Peng Zhang. 2021. Data Engagement Reconsidered: A Study of Automatic Stress Tracking Technology in Use. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 535, 13 pages. doi:10.1145/3411764.3445763
- [4] Daniel A. Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M. Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, Payam Dowlatyari, Craig Hilby, Sazed Sultana, Elizabeth V. Eike, and Yunan Chen. 2020. Mapping and Taking Stock of the Personal Informatics Literature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (Dec. 2020), 126:1–126:38. doi:10.1145/3432231
- [5] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does gamification work?—a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences*. Ieee, 3025–3034.
- [6] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-Based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). Association for Computing Machinery, New York, NY, USA, 557–566. doi:10.1145/1753326.1753409
- [7] Elizabeth L. Murnane, Tara G. Walker, Beck Tench, Stephen Volda, and Jaime Snyder. 2018. Personal Informatics in Interpersonal Contexts: Towards the Design of Technology that Supports the Social Ecologies of Long-Term Mental Health Management. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 127 (Nov. 2018), 27 pages. doi:10.1145/3274396
- [8] Harri Oinas-Kukkonen and Marja Harjumaa. 2008. Towards deeper understanding of persuasion in software and information systems. In *First international conference on advances in computer-human interaction*. IEEE, 200–205.
- [9] Yaacov Trope and Nira Liberman. 2010. Construal-Level Theory of Psychological Distance. *Psychological Review* 117, 2 (2010), 440–463. doi:10.1037/a0018963
- [10] Daphne van Zandvoort, Marloes Vredenburg, and Marit Bentvelzen. 2025. Unhealthy Comparisons to Promote Healthy Behavior? Exploring the Impact of Social Comparison Strategies in Personal Informatics. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '25). Association for Computing Machinery, New York, NY, USA. doi:10.1145/3706598.3713737