MASTER

Self-tracking data for professional use
trust and transparency in the automated interpretation of health data

van den Driesschen, T.

Award date:
2019

Self-Tracking Data for Professional Use

Trust and transparency in the automated interpretation of health data

By Tim van den Driesschen
Identity number: 1028207

In partial fulfilment of the requirements for the degree of
Master of Science
In Human Technology Interaction

Date: 22-08-2019

Version: 1.0

Supervisors: dr.ir. M.C. Willemsen & H. Rutjes, MSc

Abstract

With the rise of wearable health technology, there is an increasing interest in the opportunities provided by the captured patient-generated data (PGD). With chronic diseases as the main cause of mortality and morbidity within the European Union, this self-captured data could provide health coaches with valuable information regarding chronic disease management.

Decision support systems (DSS) using PGD offer great potential for health coaches to fit PGD in their workflow and automate the analysis of PGD to prevent data overload. However, it is important to take trust of the health coaches in the automated analysis of PGD into account. Therefore, this study aims to understand the transparency needs of health coaches in DSS that use PGD.

Through a between subject design, the effect of procedural and outcome transparency on the trust of health coaches with different levels of work experience was tested. The health coaches interacted with a DSS that visualized PGD and offered a recommendation for a coaching advice. The DSS demonstrated different levels of transparency in the recommendation through explanations.

The study found that, on the one hand, more experienced health coaches trust recommendations from DSS more when explanations about the outcome are provided compared to explanations about the procedure. On the other hand, less experienced health coaches trust recommendations from DSS more when procedural explanations are provided compared to outcome explanations. While findings of earlier research are supported, the study's unique focus on health coaches provides a new perspective on the implementation of DSS using PGD, and insights on the design of DSS with regards to transparency.

Keywords: Trust, Transparency, Automation, Decision support systems, Patient-generated data, Health coach, Work Experience

Contents

List of Figures

List of Tables

Introduction

Chronic diseases are the main cause of mortality and morbidity within the European Union (Dejonghe, Becker, Froboese, and Schaller, 2017). As chronic diseases are to a large extent preventable, health coaching has emerged in recent years as a promising lifestyle intervention and shown positive effects on health outcomes. Health coaches encourage individuals to improve their lifestyles, focus on better nutrition and more exercise, quit bad habits, and deal with their chronic conditions.

Nowadays, many devices in the consumer market, such as fitness trackers, smartwatches, and smartphones, allow individuals to capture data on their health day and night. This lead to an increasing interest in the opportunities provided by this so-called patient-generated data (PGD) (Chung, Cook, Bales, Zia, and Munson, 2015; Shapiro, Johnston, Wald, and Mon, 2012). The data captured by these wearable health devices allows for a window into the daily life of the client, which provides the health coach with valuable information for chronic disease management about the lifestyle choices and habits of their client.

Prior research mainly focuses on the opportunities of PGD within the clinical healthcare setting. Clinical decision support systems (CDSS) using PGD and electronic health records as sources of information are mentioned as promising technologies to help healthcare providers improve care quality (Chaudhry et al., 2006). CDSS analyze the often vast amounts of digital patient data and use this information to help healthcare providers effectively and reliably with decision-making, formulating a diagnosis, deliver and monitor personalized treatment plans and facilitate the relationship with the patient (Chaudhry et al., 2006; Chung et al., 2016).

These technologies also offer great opportunities outside the clinical healthcare setting, but research remains limited. Lifestyle changes needed to manage or prevent chronic diseases are to a large extent outside of clinical control (Gray, Arlinghaus, and Johnston, 2018). Clients usually need to alter their diet and exercise habits and therefore data from wearable health devices are specifically valuable to health coaches.

Potential of decision support systems and patient-generated data for health coaches

Limited research has been conducted regarding decision support systems (DSS) using PGD outside of the clinical health sector. However, the findings of studies within the clinical healthcare sector discussed below can also to a large extent be applied to health coaches as there are many similarities between the professions regarding workflow and relationship with the patient and client.

Currently, healthcare providers have consultations with patients to listen to them and discover their problems. During these consultations, healthcare providers learn essential information about the daily lives of patients. However, the time during these consultations is often limited, which makes it complicated to understand the patient.

Moreover, previous research shows that patients often find it hard to summarize experiences over longer times (Brusco and Watts, 2015). Besides, patients do not report the average of events but rather how it was at its peak and the end of an event, known as the peak-end rule (Kahneman, Fredrickson, Schreiber, and Redelmeier, 1993). Also, other biases come into play when recalling past events such as the mood-congruent memory bias (Matt, Vázquez, and Campbell, 1992). This bias causes patients to remember moments which are congruent with their current mood better than incongruent moments.

The data captured by self-tracking devices allows for a window into the daily lives of patients, which is sometimes difficult for patients to explain in a short time during a consultation (Neff and Nafus, 2016). CDSS provide a great opportunity to support healthcare providers effectively and reliably with decision-making under the given time constraints. CDSS can be used to diagnose patients as well as deliver personalized treatment plans suited to the patient's needs and preferences. By combining data from both electronic health records and PGD the quality of decision-making of healthcare providers can be improved (Chaudhry et al., 2006).

PGD offers a great source of information for CDSS (Sittig et al., 2008). Healthcare providers have already recognized the value of PGD to help achieve several goals such as supporting a diagnosis, personalizing treatment, increasing motivation and accountability, learning about patients, and facilitating discussions (Chung et al., 2015; Huba, and Zhang, 2012; Rutjes, Willemsen, Van Kollenburg, Bogers, and IJsselsteijn, 2017).

The use of PGD helps patients be better understood by their healthcare provider. PGD helps them to learn about correlations and associations between symptoms and triggers and gives empirical support for their anecdotes (Chung et al., 2015). In addition, PGD can be a starting point for more engaging conversations between healthcare providers and patients (Chung et al., 2016). Irregularities and ambiguity in PGD are conversational starting points that help healthcare providers better understand patients. Moreover, it provides opportunities for personalized treatments and can enhance the healthcare provider-patient relationship (Chung et al., 2016). The healthcare provider-patient relationship has proven to be a crucial factor in improved treatment outcome (O'Broin and Palmer, 2006; Stewart, 1995). Therefore, PGD can be an essential tool to increase the healthcare provider-patient relationship and thus, treatment outcome.

Prior research already demonstrates the value of CDSS using PGD. For example, Lambert et al. (2006) investigated in a meta-study the effect of feedback to therapists through a CDSS on patients' progression through time. The feedback allowed therapists to adjust treatment plans if the patient was at risk for treatment failure. The study found that there were small improvements for patients that were already recovering. However, the improvements of patients that were not recovering as expected were much more prominent. Hence, these findings provide evidence for the value of CDSS and PGD in healthcare and demonstrate that there are actual benefits with regards to improved treatment outcome.

In addition, Rutjes, Willemsen, and IJsselsteijn (2019a) found during interviews that health coaches see the benefits of PGD in health coaching. PGD from training sessions and daily life is seen as an important source of information. Furthermore, the health coaches appreciate the facts and figures PGD can provide. These findings again demonstrate the value of PGD in a healthcare setting.

Moreover, Chaudhry et al. (2006) reviewed studies regarding electronic health records and CDSS to search for evidence for the effects of health information technology on quality, efficiency, and cost of healthcare. The study found that the implementation of health information technology could yield real benefits with regards to improve the delivery and quality of care. However, the researchers note that in terms of interoperability of different systems and integration of consumer health technologies, a lot of work still has to be done.

All in all, prior research demonstrates that CDSS and PGD can support healthcare providers effectively and reliably with decision-making, formulate a diagnosis, deliver and monitor personalized treatment plans and facilitate the relationship between the healthcare provider and patient. Similar conditions and circumstances also apply to health coaches and their clients. More specifically, information on the lifestyle choices and habits of the client which can be obtained from PGD is valuable information in aiding chronic disease management. Therefore, the potential of CDSS and PGD for healthcare providers are also applicable to health coaches. Hence, it is important to understand what barriers are withholding health coaches from using such technologies.

Current barriers of implementation of decision support systems and patient-generated data by health coaches

While there is evidence for the great potential of CDSS using PGD in the healthcare sector, the technologies are not widely adopted in practice. Several barriers hold back the proliferation of such systems, such as data overload and a changing workflow. This is important to understand as similar barriers might withhold health coaches from adopting DSS using PGD.

Data overload: interpretation and volume

Data overload is one of the main barriers healthcare providers face when using PGD (Chung et al., 2016; Sittig et al., 2008; West, Van Kleek, Giordano, Weal, and Shadbolt, 2018). PGD is often a vast amount of digital data and is both time consuming and difficult for healthcare providers to analyze and understand (West et al., 2018). Healthcare providers also have limited time for reviewing this data during consultations (Chung et al., 2016) and often mention that PGD can be distracting during these patient visits (West et al., 2018). It is especially challenging to transform these large amounts of data into knowledge (Sittig et al., 2008) and extract the most relevant data (West et al., 2018).

PGD data can also be ambiguous, because the meaning of data can be subjective (West et al., 2018). This means that the data in itself is often not sufficient to understand a patient because it is not always clear what is healthy for the individual in question (Van Dijk, IJsselsteijn, and Westerink, 2016). This makes the interpretation of health data even more challenging. However, research suggests that it is still crucial to take advantage of the benefits of PGD (West et al., 2018; Zhu, Colgan, Reddy, and Choe, 2017).

Changing workflow: social aspects and different expectations

It is important to acknowledge that implementing CDSS and PGD will affect the workflow of the healthcare provider. Therefore, it is important to understand the requirements and needs of the healthcare provider in order to prevent resistance to change. To promote the adoption and effective use of these systems, understanding the needs of the healthcare provider and incorporating user feedback are critical (Deokar and Sarnikar, 2014). However, research on the needs of the healthcare provider lack meaning more empirical research is required (Rutjes, Willemsen, and IJselsteijn, 2019a).

Furthermore, the social aspect of change in the workflow needs to be considered (Sittig et al., 2008). Healthcare providers have different expectations of what to do with PGD than patients (Chung et al., 2016). Healthcare providers want to see trends in PGD, whereas patients wish to review PGD and use it to see anomalies extensively. The discrepancies between their expectations cause tensions between healthcare providers and patients (Zhu et al., 2017). These tensions are a significant barrier for incorporating PGD and CDSS into the workflow of a healthcare provider.

All in all, DSS using PGD offer great potential for both healthcare providers and health coaches. Furthermore, data overload and changing workflow could also withhold health coaches from implementing PGD. As demonstrated, previous research focuses mostly on healthcare providers. Therefore, this study will provide a new perspective on DSS using PGD by focusing on health coaches. With an increasing interest in health coaching, and the great potential for DSS using PGD to support chronic disease management, more empirical research regarding possible barriers in adopting these systems is highly relevant. Hence, this study will examine the relationship between the transparency of DSS and trust by health coaches within these systems.

Automation of decision support systems

Data overload remains one of the most significant barriers withholding the implementation of PGD as the analysis of PGD can be difficult and time-consuming. Automation of the interpretation of PGD in DSS could potentially prove to be a solution regarding this issue. It would allow health coaches to use PGD without the associated time constraints. However, it is not yet clear what part of the interpretation or analysis of PGD should be automated and to what extent.

In the year 2000, the important question of "which system functions should be automated and to what extent?" was already raised (Parasuraman, Sheridan, and Wickens, 2000). This question is still relevant as on the one hand, previous research shows that data overload increases the need for automation (Sittig et al., 2008; West et al., 2018). On the other hand, previous research demonstrates that healthcare providers are skeptical about the automated interpretation of data. Healthcare providers want to stay in the loop and understand how data is analyzed (Rutjes et al., 2019a). Hence, transparency in the underlying mechanism that interprets the data could potentially facilitate healthcare providers trust in the system.

Furthermore, humans can understand other humans, to be empathetic and understand the context and thus need to understand how the interpretation of the data is achieved (Fayyad, Piatetsky-Shapiro, and Smyth, 1996). This is also true for the health coaches and their clients because there are aspects that can not be captured by data alone (Van Dijk, IJsselsteijn, and Westerink, 2016). For example, a client has personal goals or progression which cannot be taken into account of the coaching advice by the automated system.

Parasuraman et al. (2000) outlined ten levels of automation based on the amount of control the human still has over the system (see Figure 1). In levels one through five, the human still has overall control over the task. The increasing levels indicate greater control of the system. In levels six through ten, the system performs the task independently. The researchers also identified four stages where automation can be applied. A particular system can have automation in all four stages at different levels. This means that for every stage one out of the ten levels of automation can be chosen. The first stage is the data acquisition stage. The second stage is the data analysis stage. The third stage is the decision-selection stage, and the final stage is action-implementation.

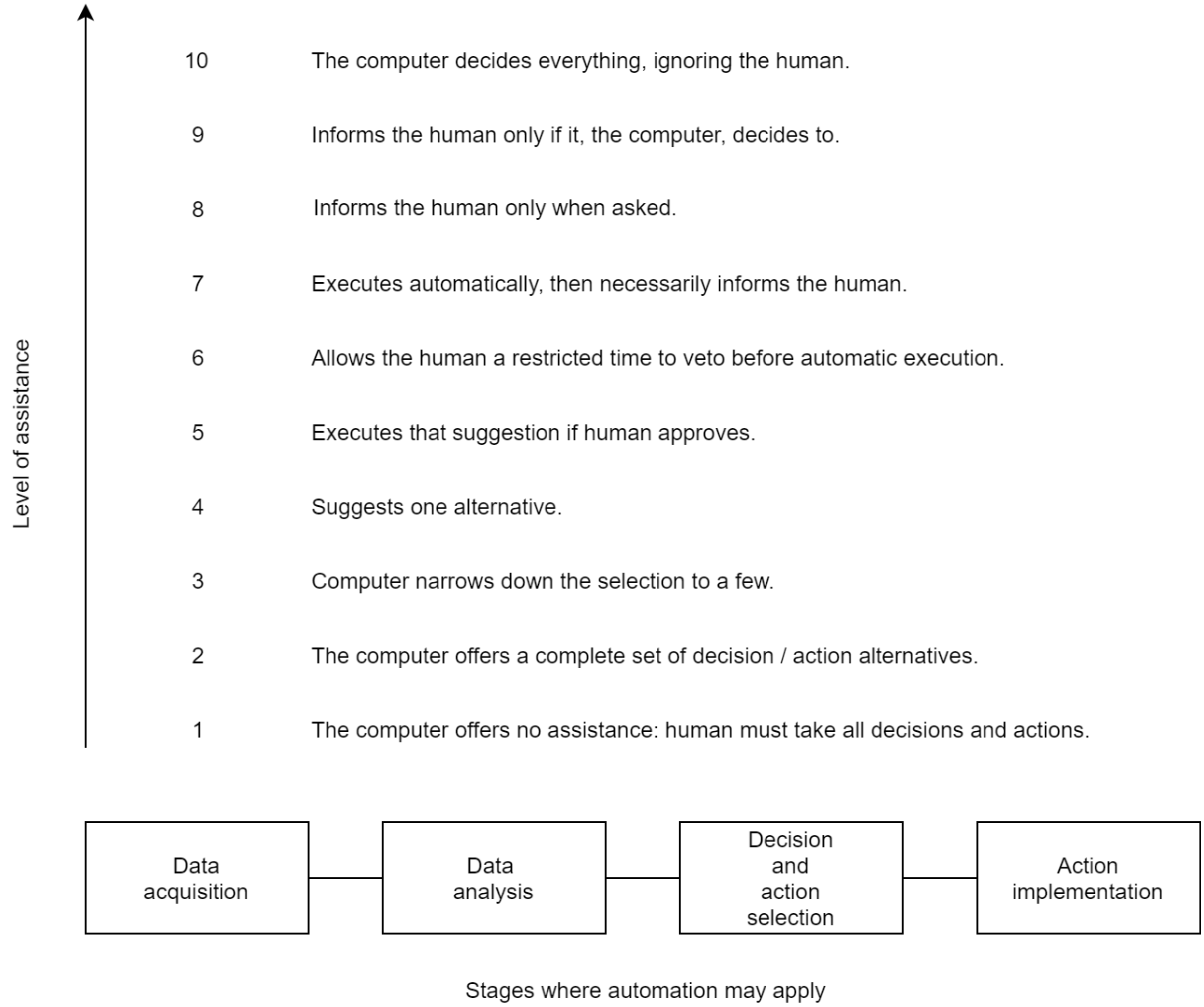

Figure 1. The Ten Levels of Automation and where they may apply. Adapted from: Parasuraman et al. (2000).

Moreover, this framework aids in choosing the right level of automation by proposing primary and secondary evaluative criteria. The primary evaluative criteria entail the consequences of the stage and level of automation for the human operator in the resulting system. The secondary evaluative criteria assess automation reliability and the costs of decision and action consequences against the stage and level of automation. Human trust in the system and use of the system is for example highly influenced by automation reliability. Reliability increases the operator’s trust whereas unreliability lowers trust and can potentially undermine the benefits of automation.

To apply the concept of Parasuraman et al. (2000) to DSS using PGD and health coaches, the data acquisition and data analysis stage are addressed first. Health coaches need to stay in the loop and understand how decisions are made (Rutjes et al., 2019a). Lack of trust in the accuracy and analysis of the data of the automated system is often mentioned as an obstacle in why healthcare professionals are not widely using PGD (Adadi and

Berrada, 2018; Alsos, Das, and Svanæs, 2012; Gagnon, Ngangue, Payne-Gagnon, and Desmartis, 2015). If health coaches want to stay in control, this means that automation at the data analysis stage should not exceed level five (Adadi and Berrada, 2018; Rutjes et al., 2019a).

However, the ideal level of automation at the data analysis stage for DSS using PGD is difficult to determine as it is unclear at what level health coaches are willing to trust these systems. As Parasuraman et al. (2000) suggest, automation reliability influences trust of the health coach in the DSS. However, more research is needed to understand the dynamic interplay between the level of transparency of the system and the level of trust of the health coach in the automation of the interpretation of PGD in a DSS.

Following the concept of Parasuraman et al. (2000), the data acquisition stage has been automated within the sector of health coaching, specifically for self-trackers. Signals from sensors are already translated to interpretable data. For example, movements from wearables are translated into steps. In the data analysis stage, automated summarizations, visualizations, and analysis of the data could assist the health coaches with having a better understanding of the daily life of the patient. Accordingly, it can provide recommendations to the health coaches. Based on previous literature and the current automation of the data acquisition stage by wearable technology devices, this study will focus on the automation of the data analysis stage.

Overall, automation reduces time and increases effectiveness by analyzing and interpreting PGD. The automation could be a solution for the data overload barrier health coaches face when using PGD. However, the trust health coaches have in the automated analysis of PGD should be taken into account. Therefore, it is important to understand what trust is and what factors influence trust in automated systems.

Trust

Barriers that are seen in the healthcare sector could potentially also withhold health coaches from utilizing PGD. It is important to recognize these potential barriers and find solutions to overcome them in order to reap the benefits of PGD.

A solution to data overload can be the automated analysis of PGD in a DSS. However, when automating PGD new challenges will arise such as trust issues regarding the analysis. To health coaches it is important to stay in the loop and understand how the analysis is performed (Rutjes et al., 2019a).

In order to understand how trust issues affect the willingness of health coaches to utilize and rely on DSS, it is important to first understand what trust is and how to define it.

Trust has traditionally been challenging to define as it is a multidimensional concept (Mayer, Davis, and Schoorman, 1995; Rousseau, Sitkin, Burt, and Camerer, 1998). McKnight, Choudhury, and Kacmar (2002), have validated measures of trust in several dimensions of trust. The demensions they identified were: dispositional trust, institution based trust, trusting beliefs, trusting intentions. The two main dimensions of interest for this study are trusting beliefs and trusting intention.

Trusting beliefs means in the context of this study, the health coach's perception that DSS have attributes that are beneficial to the health coach. The DSS must have the competence (the ability of DSS to do what the health coach needs), benevolence (DSS working in the best interest of the health coach), and integrity (DSS report results honestly) to create trust (McKnight et al., 2002). To measure this dimension of trust, questionnaires can be used.

The trusting intention, is defined within the context of this study as the willingness to depend on the recommendations of DSS by a health coach. The willingness to depend on the recommendation can be measured with the compliance of the health coach with the recommendation In this study, both trusting beliefs and willingness to depend will be used as measures of trust and will be referred to as *trust*.

Trust of health coaches in decision support systems

Data overload is a barrier withholding healthcare providers from utilizing PGD. This barrier could potentially also apply in the context of health coaches. This study proposes the solution of automation to help health coaches overcome the potential data overload barrier. However, automation itself has challenges that need to be overcome. Mainly challenges regarding trust health coaches have in the automated interpretation of PGD in DSS. Therefore, it is important to understand what factors influence trust in automated systems.

Trust in automated systems is essential before humans are willing to operate such systems (Adadi and Berrada, 2018; Gagnon et al., 2015; Hoff and Bashir, 2014; Parra and Brusilovsky, 2015). This need for trust is even more significant in healthcare-related settings where misuse of automated systems can have significant consequences for the patient outcome (Adadi and Berrada, 2018; Gagnon et al., 2015; Hoff and Bashir, 2014). Lack of trust in the accuracy and analysis of the data of the automated system is often mentioned as an obstacle in why healthcare professionals are not widely using PGD (Adadi and Berrada, 2018; Alsos, Das, and Svanæs, 2012; Gagnon et al., 2015; Kellermann and Jones, 2013; Nunes and Jannach, 2017).

Hoff and Bashir (2014) created a three-layered framework for conceptualizing trust (*Figure 2*). This framework can be used to discover important factors in increasing trust of the health coach in DSS.

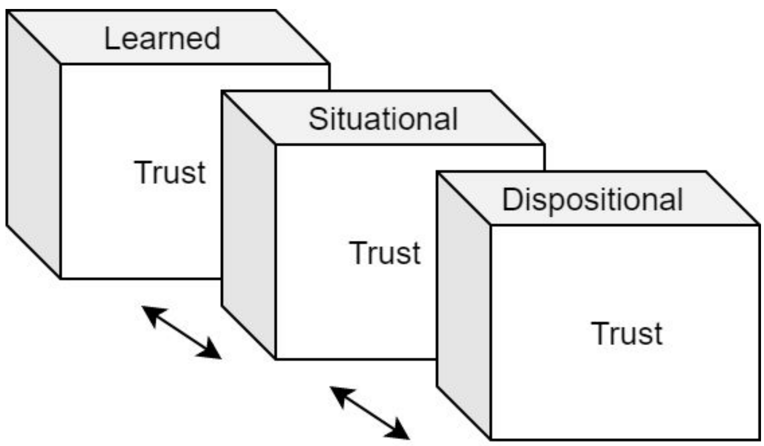

Figure 2. The Three-layered Framework for Conceptualizing Trust Variability in Automated Systems. Adapted from: Hoff and Bashir (2014).

The first layer, the dispositional layer, is about long-term tendencies in individuals to trust automation and is not dependent on the context or the system that is used. Therefore, factors in the dispositional layer are not changed easily. A health coach's beliefs shape trust in the dispositional layer. The four primary sources of dispositional trust are culture, age, gender, and personality. The researchers found that older people have more trust in tools used for decision-making than younger people. Gender, on the other hand, does not have a particular direction but differences often do exist. All in all, this shows there are great individual differences between individuals.

The second layer is the situational layer and consists of two primary sources of variability, the internal and external variability. Internal variability is found to originate from self-confidence, subject matter expertise, mood, and attentional capacity. External variability, on the other hand, arises from the type of system that is used, system complexity, system feedback and transparency, and workload. The researchers mainly emphasize the importance of system feedback and transparency. These findings thus show that for the development of trust of the health coach in DSS, it is important that the system is transparent and that the DSS fit into the workflow of the health coach.

The third and final layer is the layer of learned trust. Learned trust can be divided between initially learned and dynamically learned trust. Health coaches use knowledge from past interactions with automation when assessing the trustworthiness of a new system. This prior knowledge is a form of initially learned trust. Moreover, during an interaction, an automated system may perform inconsistently, and trust is likely to fluctuate similarly. This is called the dynamically learned trust.

Based on the three-layered framework of Hoff and Bashir (2014), the situational trust layer needs to be taken into account in the development of DSS using PGD. Influence can be exerted on the factors in the situational layer such as the system feedback and transparency. This means that these factors are crucial in the development of trust of health coaches in the system.

In line with Hoff and Bashir (2014), other studies found similar results that the main factors in the development of trust can be distinguished in human-related and automation-related factors (Schaefer, Chen, Szalma, and Hancock, 2016). The human-related factors again indicate that there are individual differences between health coaches in trusting DSS.

Besides, it was found that transparency in the automated system leads to higher levels of trust in an automated system (Gagnon et al., 2015; Schaefer et al., 2016). Also, providing transparency in the form of explanations increases the trust users have in automated systems (Cheng et al., 2019; Kilzelcec, 2016).

These studies show that transparency in an automated system is essential for health coaches to develop trust and therefore, crucial for their willingness to use such systems. So in order to utilize the potential of PGD in DSS through the automated analysis of PGD, health coaches need to trust the DSS. Transparency is a solution to the challenges regarding trust in automated systems. Therefore, the possibilities of transparency will further be investigated.

Transparency of decision-support systems through explanations

Transparency can be a possible solution for the challenges regarding trust. Transparency can give health coaches insights into the inner workings of a DSS and thus facilitate the feeling of still being in the loop. Therefore, the following section investigates transparency in automated systems and how this may apply to health coaches.

Transparency in DSS can be provided in several different ways. A possible way of providing transparency in the analysis of PGD by a DSS is through visualizations (Lee, Garrity, and Newman, 2019). Visualization of the data used to formulate the coaching advice can increase the transparency of the DSS by giving insights into the data that was used. Therefore, it is important to critically consider the visualization and analysis of DSS in promoting the adoption and effective use of DSS using PGD. Visualization techniques have shown that complex data can easily be understood by properly visualizing it (Fekete, Van Wijk, Stasko, and North, 2008; Saket et al., 2018). In this way, new patterns can be found, and previously unknown relationships can be discovered (Claes and Moere, 2015; Saket et al., 2018; Thudt, Hinrichs, and Carpendale, 2012). The visualizations can help health coaches with their desire to see trends in PGD (Zhu et al., 2017) and thus help with the adoption of DSS and PGD in the workflow of health coaches.

While visualization techniques help to increase transparency to a certain extent, they do not explain the inner workings of the DSS of how the data is analyzed and the coaching advice is created. In other fields, increasing transparency in an automated system by explaining the inner workings generally builds trust (Beck, Dzindolet, and Pierce, 2007; Cheng et al., 2019; Dzindolet, et. al., 2002; Kilzelcec, 2016, Lyons, et. al., 2017; Nilashi, Jannach, Ibrahim, Esfahani, and Ahmadi, 2016). Transparency through explanations of the workings of DSS is one of the fundamental drivers of trust of healthcare providers (Cheng et al., 2019; Nunes and Jannach, 2017; Schaefer et al., 2016).

However, Research on the implications of transparency of DSS using PGD on the trust health coaches have remains limited despite the high relevance of the topic (Rutjes, Willemsen, and IJsselsteijn, 2019b). Therefore, this study aims to answer the following research question:

“What are the effects of transparency on the trust health coaches have in decision support systems that use patient-generated data?”

It is expected that providing transparency through explanation on the data analysis should increase trust of health coaches in the DSS (Beck et al., 2007; Cheng et al., 2019; Dzindolet, et. al., 2002; Kilzelcec, 2016, Lyons, et. al., 2017; Nilashi et al., 2016; Nunes and Jannach, 2017; Schaefer et al., 2016). Thus, the first hypothesis is formulated:

H1: Decision support systems with transparency through explanations are trusted more than decision support systems without transparency through explanation.

In general, it is expected that transparency in the recommendation through explanations increases the trust in the DSS. However, mixed effects were found for the type of explanations that were used to provide transparency (Arnold et al., 2006; Cheng et al., 2019; Kizilcec 2016; and Wang and Benbasat 2007).

To answer the proposed research question and test the hypotheses, this study will conduct an experiment where health coaches will be offered a fictitious case. They have to work with a DSS that uses a regression model which automates the data analysis stage. The dashboard will show the health coaches a recommendation based on the PGD for coaching advice. This recommendation will be shown in written text as this is suggested to facilitate understanding (Kacprzyk and Zadrożny, 2009). Moreover, the PGD will be visualized in bar charts as this is suggested to be an intuitive way to understand PGD (Lee, Garrity, and Newman, 2019).

Overall, it is expected that transparency increases the trust health coaches have in the recommendation of DSS. Visualization provide some form of transparency and allow health coaches to inspect the PGD and look for trends and anomalies. However, more transparency in the analysis of the data needs to be provided to increase the trust of health coaches in the analysis of PGD. This transparency can be provided in the form of an explanation of the inner workings of the DSS. However, the type of explanation that health coaches need is unclear.

Expertise and the type of transparency

To fully utilize the potential of PGD in health coaching, it is important for DSS to be transparent in how they analyse the data. Explanations of the inner workings of automated systems have shown to increase the trust of users. However, mixed effects for the type of transparency was found. Therefore, to understand the needs of health coaches, different forms of explanations will be investigated.

The transparency through explanations about the inner workings of DSS can be provided in several ways. Two types of explanations can be distinguished: procedural and outcome explanations (Arnold et al., 2006; Cheng et al., 2019; Kilzelcec, 2016). Procedural transparency explains what procedural steps an automated system uses to create the recommendation. Procedural justice theory says that individuals trust recommendations more when the procedure behind an automated system is revealed (Lind and Tyler, 1988). Outcome transparency not only explains the procedural steps but also justifies the recommendation of the system. However, providing more information has shown to erode the trust for some users of automated systems (Arnold et al., 2006; Cheng et al., 2019; Kilzelcec, 2016).

Research has shown that different levels of expertise need different types of explanations for transparency before an automated systems is trusted (Arnold et al., 2006; Cheng et al., 2019; Desai et al., 2019; Hoff and Bashir, 2014; Schaffer et al., 2019). For the different levels of expertise, the distinction between laymen, novices, and experts is often made. Users of a system with no domain knowledge or work experience are referred to as laymen. Users with some domain knowledge but are new in the field, and thus have little work experience, are referred to as novices. Finally there are users with domain knowledge and work experience. These individuals are referred to as experts.

In this study, several dimensions of expertise need to be considered. First, the dimension of health coaching needs to be considered. Users without any education or training as a health coach could be considered laymen. Users that already have some domain knowledge but still lack work experience as health coaches could be considered novices. Finally, experts are users that have domain knowledge and work experience in health coaching.

Moreover, other dimensions of expertise should be considered, such as, knowledge about regression models and formulas, and experience working with self-tracking data. These forms of expertise are considered because this study employs a DSS that uses a regression model to analyze PGD and more knowledge in these dimensions can be seen as a form of expertise.

This study will aim to test its hypotheses mainly on the groups of expertise regarding health coaching and regression model and formula expertise. *Figure 3* shows a graphical representation of these groups.

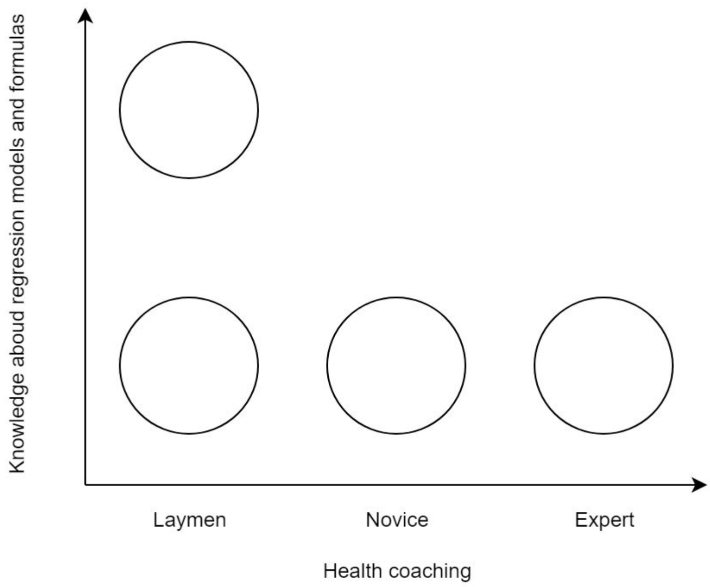

Figure 3. Levels of Expertise in Health Coaching and Regression Models and Formulas.

Prior research demonstrates that providing procedural transparency to laymen increases trust in the automated system (Eslami et al. 2019; Kizilcec, 2016; Wang and Benbasat, 2007). However, when outcome transparency was provided, the effect was nullified and was comparable to the situation where no transparency was provided. Kizilcec (2016) assigned this phenomenon to possible confusion of the layman on the outcome transparency due to the difficulty understanding the explanation. This confusion shifted away from the focus from the explanation, and thus, the outcome transparency did not increase trust.

This finding indicates that there is an optimal form of transparency when laymen use an automated system. It is thus expected that laymen health coaches trust a recommendation or coaching advice from a DSS more when explanations regarding the procedure are provided than explanations regarding the outcome. Therefore, the following hypothesis is formulated:

H2: Laymen health coaches trust recommendations of decision support systems with procedural transparency more than outcome transparency.

Novices and experts are different than laymen as they have domain knowledge. However, research has shown that novices have different needs when it comes to transparency compared to experts (Arnold et al., 2006; Cheng et al., 2019; Desai et al., 2019; Schaffer et al., 2019). Novices want more feedback on the procedure rather than the outcome because novices are still learning to work within the problem domain (Arnold et al., 2006). Procedural transparency helps novics better understand the domain by explaining how the system creates the recommendation. Schaffer et al. (2019) showed that novices trust recommendations of an automated decision most when procedural explanations are provided. Therefore, it is expected that novice health coaches prefer procedural transparency over outcome transparency. Hence, the following hypothesis is formulated:

H3: Novice health coaches trust recommendations of decision support systems with procedural transparency more than outcome transparency.

Furthermore, research has shown that experts have different transparency needs compared to novices (Arnold et al., 2006; Cheng et al., 2019; Desai et al., 2019; Schaffer et al., 2019). Experts, users who have both domain knowledge and substantial work experience, report a higher need for transparency to trust an automated system than novices (Arnold et al., 2006). It was found that experts require a domain-oriented argument before they trust and adhere to the advice from an automated system (Schaffer et al., 2019). Experts need outcome transparency as they want to understand how the system came to a given recommendation rather than the general procedural steps (Arnold et al., 2006). Finally, with the knowledge that health coaches like to stay in the loop when it comes to the interpretation of health data and they would not trust the automated analysis of health data (Rutjes et al., 2019a). It is expected that expert health coaches trust outcome transparency more than procedural transparency. Hence, the following hypothesis is formulated:

H4: Expert health coaches trust recommendations of decision support systems with outcome transparency more than procedural transparency.

As the DSS in this study will use a regression model to create the recommendation for coaching advice, knowledge about regression models and formulas may also be seen as a form of expertise. Individuals with more experience using regression models may not experience the confusion that was observed in Kizilcec (2016) and thus not have the nullifying effect in trust when seeing outcome transparency. Therefore, it is expected that individuals with more knowledge about regression models and formulas trust outcome transparency more than procedural transparency leading to the following hypothesis:

H5: Users with more knowledge of regression models and formulas trust decision support systems with outcome transparency more than procedural transparency.

This study aims to investigate the different needs with regards to transparency in DSS that automate the interpretations of PGD for health coaches. These findings from the hypotheses will provide a better insight on how to design a DSS that are trusted by health coaches with different levels of expertise. In this way, this study hopes to encourage the use of PGD and take advantage of its great potential for improving health coaching.

Method

An experiment was designed in order to study what the effects of transparency are on the trust health coaches have in decision support systems (DSS) that use patient-generated data (PGD). A data set from a fictitious client was shown to the health coaches, the participants of this study. The software program Tableau Public was used to create a dashboard that visualized the data from the client and enabled the participants to interact with it. In addition, the dashboard offered the participants a recommendation for a coaching advice to the clients. The participants could decide whether to give this recommendation to the client or not. The recommendation offered to the participants was in every situation the same. However, the explanation on how the recommendation came about differed in levels of transparency.

Design

A 2 (procedural transparency, outcome transparency) x 3 (work experience: layman, novice, and expert) design was used. This design used a partly within-subject design as all participants would first see a low transparency condition and afterwards either the procedural transparency or the outcome transparency condition. The main comparison of interest was the difference between trust in the procedural and outcome transparency after correcting for the low transparency condition. Work experience is a between-subject factor because work experience is an inherent property of the participant.

Participants

Different participants were chosen to join the experiment based on their work experience in health coaching. Different groups were asked to participate in the experiment: experts in health coaching, novices in health coaching, laymen in health coaching, and students with knowledge about regression models and formulas. The experts were recruited either via the alumni network from the Fontys Sporthogeschool or via recruitment through emails to companies specialized in health coaching. The experts worked as health coaches (lifestyle coaches, personal trainers, dieticians and exercise physiologists). The novice health coaches that were recruited were students from the Fontys Sporthogeschool and were in the third year of their bachelor. The students had minimal experience working with clients. They only gained experience from an internship during their study. The laymen health coaches were recruited via the University of Technology in Eindhoven and followed either the course Behavioral Research methods (second year course of Bachelor Psychology and Technology) or Stochastic Processes (second year course of Bachelor of Applied Mathematics). The layman health coaches were recruited based on their knowledge of regression models and formulas. Therefore, the layman health coaches were novices in working with statistical models, such as regressions.

The experts participated voluntarily and performed the experiment in their own time. Therefore, the experts were compensated with a ten euro gift certificate. Both the Fontys and TU/e students participated voluntarily as part of a course. They were compensated by entering a raffle with one in three chance of winning a ten euro gift certificate. In total, 111 participants participated in the experiment of which 61 laymen, 33 novices, and 17 experts.

Setting, Materials, and Stimulus

Setting

An online survey was used to execute the experiment. The experiment had to be performed on a laptop or PC. The experts had the opportunity to participate in the experiment in their own time. The Fontys and TU/e students were given the opportunity to do the experiment during lectures. The participants were asked to perform the experiment individually. The experiment took approximately 20 minutes to complete. The researchers were present when the experiment was performed with the students following the course Stochastic Processes of Fontys and the TU/e.

Materials

A dataset from an open database was used as an inspiration to create the case of the client (Cecaj, 2016). This dataset was slightly altered to assure that the recommendation was the best solution to help the client. The dataset contained sleep data (minutes asleep, minutes awake, and minutes in bed) and movement data (steps, minutes sedentary, minutes light movement, and minutes of exercise). Moreover, coffee and alcohol consumption were added to the dataset.

A fictitious case of a client named Sam was created. The client used a self-tracker to capture data on different health-related factors throughout the day. The factors included sleep time, time awake at night, light intensive movement, time exercise, time being sedentary, steps taken, cups of coffee drank, and alcohol consumptions drank. These factors were chosen based on the effects they have on sleep quality ("Deep Sleep: How to Get More of it – American Sleep Association," 2019; "Slaapproblemen | Thuisarts," n.d.), and the ability to track them with a self-tracker. Furthermore, a sleep-score factor was also given. This factor was the perceived quality of sleep by the client, Sam, and was made up.

Tableau Public was used to create the dashboard that visualized the client's data and enabled participants to look through the data (*Figure 4*). Tableau Public allowed to make the visualizations in an intuitive interactive way using bar charts.

Limesurvey was used for creating the online environment in which the experiment was performed. Limesurvey allowed for the dashboard to be embedded in the experiment together with the questionnaires.

Stimulus

The dashboard

The coaches were asked to imagine the client Sam. Sam would be visiting their practice the next day for sleep problems and tiredness throughout the day. The coaches were sent the data including a recommendation for health advice before the meeting with the client Sam. The coaches were asked to review the data and recommendation, and formulate an advice to Sam based on this.

The interactive dashboard visualized the data and recommendation for the coaches (*Figure 4*). The participants were able to look through the data and find relationships, correlations and associations between the different health factors. The recommendation was the automated interpretation of the PGD by the DSS. The recommendation offered the participants an advice they could decide to give to the client Sam or not.

This setting was chosen because the setting would fit into the workflow of health coaches and allowed them to go through the data. The fitting of PGD into the workflow was one of the main barriers withholding health coaches from utilizing it and therefore, cannot be ignored.

Figure 4. The Dashboard with Procedural Transparency. Showing: the Recommendation together with the procedural transparency, the Sleep times, the Sleep Score, the Coffee and Alcohol Consumptions, and all Movement data from the Client.

The manipulation

Three different levels of transparency (low, procedural and outcome transparency) were used. They differed in the way in how they explained the recommendation came about. The low transparency condition (*Table 1*) only showed the recommendation together with the visualized data. Visualizations can already be seen as a low form of transparency. Visualizations do not explain the recommendation from the CDSS but allow for them to be confirmed.

The other levels of transparency additionally explained on what basis the recommendation was created. The procedural transparency condition (*Table 1*) explained the procedural steps that were taken to generate the recommendation. It explained the recommendation was based on the lifestyle factors that influence the sleep of a large group of people. The outcome transparency condition also showed the regression formula that was used to create the recommendation (*Table 1*). The client data was slightly altered so that exercise was the factor with the most significant impact on the time of sleep. Similar operalizations of procedural and outcome transparency were used by Kizilcec (2016). The complete dashboard for each condition can be found in *Appendix B*. There are no differences in the data for each condition, only a different transparency condition is shown.

Table 1.
The Manipulations in the three Transparency Conditions.

Transparency	Recommendation
Low	‘The tool generates the following recommendation for Sam: Exercise has the most positive influence on how many hours Sam sleeps. Thus, advice regarding exercise has the greatest potential to improve his sleep.’
Procedural	‘The tool generates the following recommendation for Sam: Exercise has the most positive influence on how many hours Sam sleeps. Thus, advice regarding exercise has the greatest potential to improve his sleep. **Explanation of the recommendation** Based on data from a large group of people, the influence of lifestyle factors (such as less use of coffee, more exercise or more time in bed) on sleep duration was determined The tool has calculated this for Sam. It has been found that if Sam starts exercising more, this probably has the largest effect on his sleep duration.’
Outcome	‘The tool generates the following recommendation for Sam: Exercise has the most positive influence on how many hours Sam sleeps. Thus, advice regarding exercise has the greatest chance of improving his hours of sleep. **Explanation of the recommendation** Based on data from a large group of people, the influence of lifestyle factors (such as less use of coffee, more exercise or more time in bed) on sleep duration was determined. This is shown in the following formula: Sleep [hours] = Time in bed [hours] -1 - 0.06 x Coffee [number of cups] - 0.09 x Alcohol [number of glasses] + Sport [hours] + 0.2 x Light movement [hours] The tool has calculated this for Sam. It has been found that if Sam starts exercising more, this probably has the largest effect on his sleep duration.’

Measurements

The first dependent variable that was measured in this experiment was Trust. Trust was measured using a questionnaire and by the compliance of the participant with the recommendation. The questionnaire contained six questions which were questions from already validated questionnaires (Kizilcec, 2016; Lyons, 2017; and Schaffer et al., 2019). The questionnaires were adapted to fit this particular experiment. Afterwards, The questionnaire was evaluated with the trusting beliefs construct from McKnight et al., (2002). The questions were answered on a 7-point Likert scale ranging from totally disagree to totally agree. From the six questions, two were negatively phrased to prevent acquiescence bias.

Moreover, compliance was measured by asking the participants to select a factor of which they thought would have the most effect on the sleep of the client.

Furthermore, the participants also filled in two open questions. The first open question asked what information participants were missing in order to give good coaching advice. The second open question allowed participants to provide general remarks about the tool. The final measurements that were taken were age, sex, years of work experience as a health coach, experience with data from self-trackers, and knowledge about formulas and regressions.

To measure the knowledge of regression models and formulas, three questions were used. The first: *"The tool that you just viewed connected lifestyle factors with clients' issues through a regression model. To what extent are you familiar with regression models?"*. The second: "*How many courses did you follow in which regression was discussed? (For example, mathematics or statistics courses in high school or at university.)*". The third: "*In the regression model below, can you indicate which predictor (A or B) has the most influence on the outcome of Y? Y = 5 + 3A + 2B + e (Note: if you don't know the answer, don't guess, but enter "I don't know")*". These questions were designed to measure the knowledge of the participants with regards to formulas and regressions. All the questions with the answer options can be found in *Appendix A.*

Procedure

At first, the participants were greeted with a welcome text and the informed consent form. When they agreed with the terms and conditions, they were allowed to start the experiment.
Next, the participants were presented with the dashboard and the low transparency condition. Here they were asked to carefully look at the data and the recommendation and formulate a piece of advice for the client with regards to the clients sleeping problems.
Next, they had to choose one of the factors that was tracked by the client on which they based their advice to measure the compliance. After that, they were asked to elaborate on their advice in a few sentences. Afterwards, they were asked to fill in a questionnaire measuring their trust in the recommendation the tool gave.
Next, the participants were given either the procedural or the outcome transparency condition. After this, they again had to formulate a piece of advice for the client. Next, the same questionnaire measuring trust was filled in again.
At the end of the experiment, the participants were asked their age, sex, work experience with clients, work experience with PGD, and their knowledge about formulas and regressions. They were also given the opportunity to provide general remarks about the

experiment. Finally, they were given the opportunity to enter an email address if they would like to receive the compensation for their participation.
The whole procedure is visualized in a schematic representation that is shown in *Figure 5*.

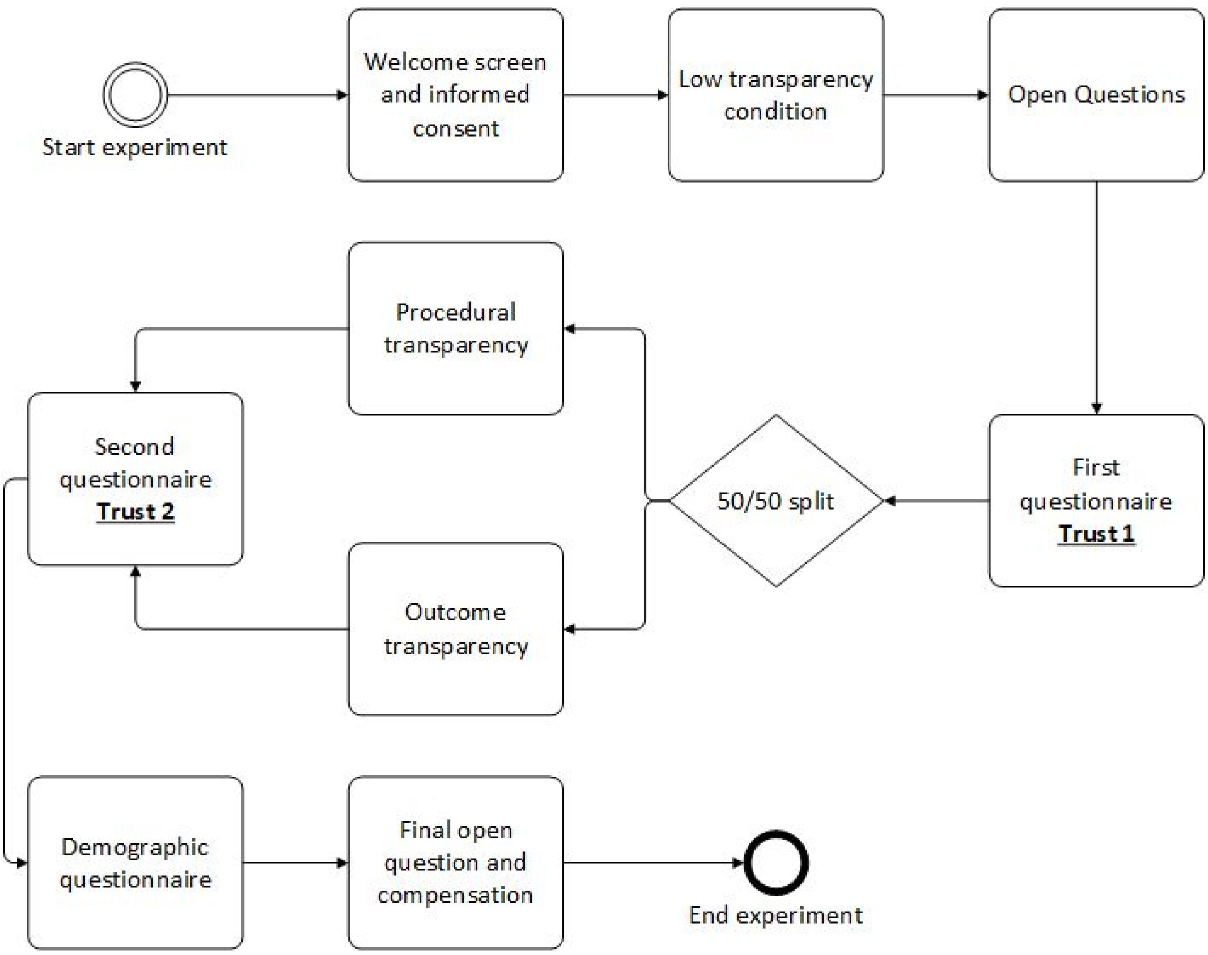

Figure 5. A Schematic Overview of the Experiment.

Statistical analysis

The analysis of the data was performed in two parts: a quantitative analysis and a qualitative analysis.

To test the hypotheses, a quantitative analysis was performed. First, the questionnaires measuring trust were created into scales; the Cronbach's alpha analysis was used to test the reliability of the scales. To answer the first hypothesis, a Wilcoxon signed-ranks test was performed on the trust measured after seeing the low transparency condition and the trust measure after viewing the transparency through explanation condition. This was a within participant comparison of the first trust measure and the second trust measure.

Hoff and Bashir (2014), showed there are great individual differences in the trust individuals have in automated systems. Therefore, to answer hypotheses two to five this study looked at the trust increase after seeing transparency through explanation (ΔTrust = Trust 2 -Trust 1).

A multiple regression was used to test the effects of work experience and the type of transparency and their interaction on the increase in trust between participants. The other measured variables were added in the multiple regression model as covariates. To inspect the interaction effect between the type of transparency and the Work Experience, the margins plot was used.

To test the influence of work experience, transparency and their interaction on the Compliance of the participants, a binomial logistic regression was used, with Compliance as the outcome variable. The applicable assumptions were tested for each analysis, and if they were not met, appropriate steps were taken to analyze the data.

In the experiment, the participants were given two open questions. The questions were asked after seeing the tool for the first and second time. The open questions were only examined and coded by the author. Therefore, the inter rater reliability could not be provided, and no substantial conclusions could have been drawn from these results. The qualitative analysis was performed for two main reasons. First, to investigate if the participants had all the data they needed to offer proper coaching advice, and if not what information they were missing. Secondly, to give participants the opportunity to raise any questions or voice their thoughts on the study in order to learn if any adjustments need to be made in future research regarding the experiment. To answer the first question, the data was coded. Similar codes were grouped, and the codes were examined on whether they could be solved with more information from PGD or a conversation. To answer the second question, the answers from the participants were reviewed and summarized.

Results

The study took place between the 13th of May and the 16th of June of 2019. Before the analysis, 59 participants with incomplete submissions were removed from the data. Most of these participants stopped after seeing the first page of the experiment or stopped halfway through the experiment.

Demographics

The resulting sample consisted of 111 participants of which 59 male, 50 female, and two who preferred not to say, with age between 18 and 62 years (M = 23.1 years, SD = 6.5 years) (*Figure 6*). The participants had an average of 1.4 years of working experience with clients (M = 1.39, SD = 2.41) of which 61 participants had no working experience with clients at all (*Figure 7*).

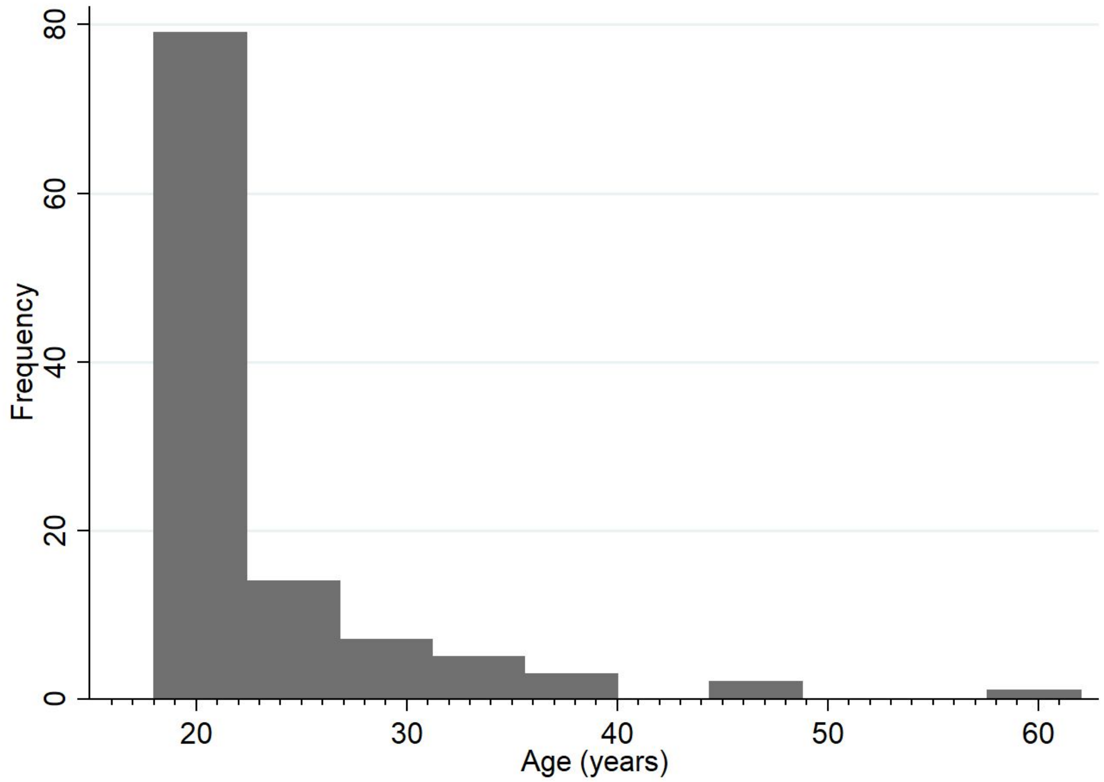

Figure 6. The Distribution of the Age (years) of the Participants.

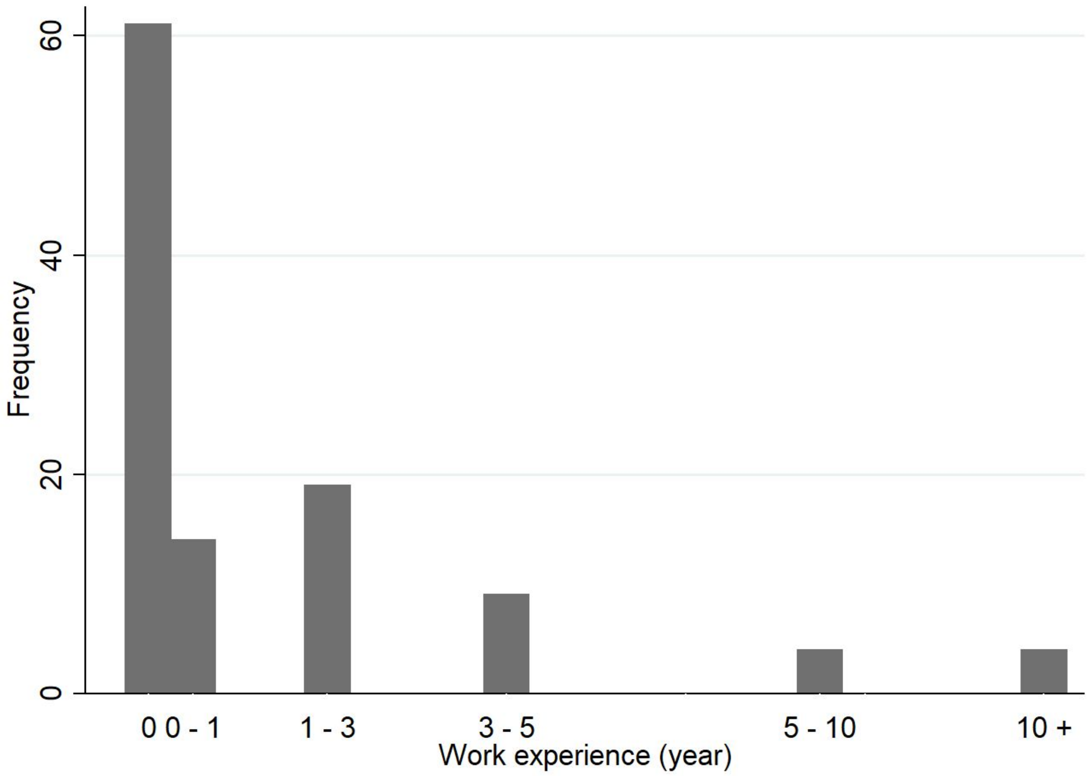

Figure 7. The Distribution of Work experience from all the participants.

Overall, the sample was skewed because the participants were young with little work experience. Participants with much work experience are underrepresented in the sample.

To inspect the distribution of the participants over their knowledge about statistical models and formulas, and work experience as a health coach, a scatter plot was used (Figure 8). The self-reported question: "*The tool that you just viewed connected lifestyle factors with clients' issues through a regression model. To what extent are you familiar with regression models?*" was used as a measure of knowledge about regression models in this plot.

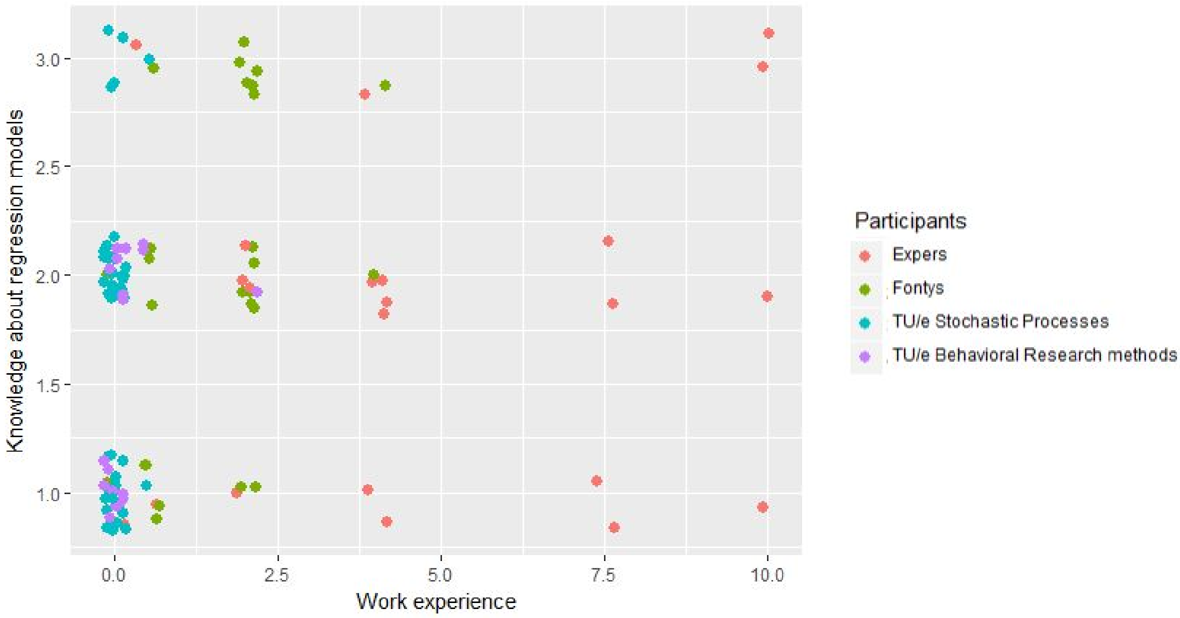

Figure 8. The distribution of Participants over Work Experience as Health Coaches (years) and their self reported Knowledge about Regression Models (Knowledge about regression models, 1: not familiar, 2: somewhat familiar, and 3: very familiar).

Figure 8 shows that most participants were not or somewhat familiar with regression models. Furthermore, the experts and novices (Fontys students) are not well defined groups in the sample. There are experts with less work experience than some Fontys students and Fontys students with more work experience as expected. Therefore, further analysis will use the years of work experience as a measure of expertise rather than the recruited group (layman, novice and expert health coaches). Laymen health care providers will be defined as having no work experience. Novices will be defined as having between zero and three years of working experience. Finally, experts will be defined as having more than three years of working experience.

Quantitative results

First, the answers to the questionnaires regarding trust were examined. In *Figure 9*, the responses to the questions are displayed. Question four and six are reversed in the graph because the questions were negatively phrased. All questions were answered on a Likert scale ranging from 1-7, where 1 = totally disagree, and 7 = totally agree.

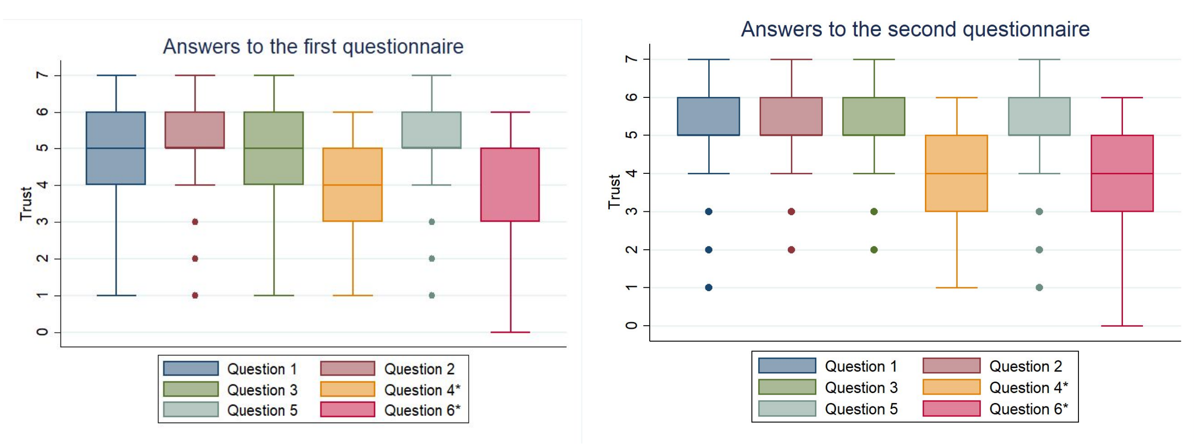

Figure 9. The Answers to the Questions measuring Trust. Left shows the Answers for the Questions after seeing the tool for the first time, right shows the Answers for the Questions after seeing the tool the second time. Question four and six were Reversed coded.

Both the first time and the second time the questionnaire was filled in by the participants, the answers score high on trust. The first time participants saw the recommendation, 49 out of 111 participants agreed with the recommendation. After seeing the recommendation for the second time, either with procedural or outcome transparency, 75 out of 111 participants agreed with the recommendation. Furthermore, even though question four and six were reversed, they still scored lower on trust compared to the positively phrased questions. Some participants may not have noticed the questions were negatively phrased.

Scales

To measure the internal consistency of the trust questionnaire, the Cronbach's alpha was used, and reliability analysis was carried out on all six items measuring trust. Cronbach's alpha showed the first questionnaire (low transparency) to reach good reliability, $\alpha = .78$. For the second questionnaire (procedural or outcome transparency), the Cronbach's alpha reached good reliability of, $\alpha = .83$ (*Table 1)*. When removing item four from both scales, the alpha's would be .81 and .85, respectively. Therefore removing this item was considered, because item four is negatively phrased (*Appendix A*) which some participants may not have noticed. However, the interpretation of the subsequent analysis did not change. Because the scale was already good, it was decided not to remove item four from the scale.

To measure the knowledge on regressions of participants, they had to answer three multiple-choice questions. These items together had a Cronbach's alpha of, $\alpha = .60$, which makes it acceptable (*Table 2*).

Table 2.
The Cronbach's alpha scores for the Scales Trust and Knowledge about Regression Models and Formulas.

Scale	Cronbach's Alpha	Mean	Std. Deviation	Min	Max	N of items
Trust1	.78	2.33	0.87	-1.17	4	6
Trust2	.83	2.32	0.91	-0.83	4.33	6
Knowledge Reg.	.60	0.37	0.60	-0.67	1.67	3

Transparency through explanations increases *Trust*

The first hypothesis, transparency provided by explanations would increase the trust of the participants in the recommendation, was tested. The assumptions for a paired sample t-test were tested before performing the analysis. The data could not be assumed to be normally distributed (Shapiro-Wilk $W = 0.94$, $p < .05$). Therefore, a Wilcoxon matched-pairs signed-ranks test was performed to test the effects of transparency on Trust.

The Wilcoxon Signed-Ranks test indicated that the trust after the recommendation with explanations as transparency (Trust2) was not statistically significantly higher than the trust after seeing the recommendation without transparency through an explanation (Trust1) ($Z = -0.741$, $p > .05$).

This result shows that transparency through explanations did not increase the trust in the recommendation of the DSS.

Regression analysis to test the effects of expertise on the trust in DSS using different types of transparency

For the following analysis, the continuous variable *Age, Regression_expert,* and *Work experience* and the categorical variable *Sex* (male, female, and prefer not to say), *Experience data* (much experience, some experience, and no experience) and *Transparency* (procedural and outcome transparency) were used as predictor variables for the difference in ΔTrust. All variables were centered, and the continuous variables were standardized.

A multiple regression model was used to test whether laymen and novices trust procedural transparency more compared to outcome transparency (H2 and H3). Moreover, this model was used to test if experts trust outcome transparency more compared to procedural transparency (H4) and if participants with more understanding of regression models trust outcome transparency more than procedural transparency (H5).

The residuals were not normally distributed, and homoscedasticity could not be assumed. Because of the skewed data in work experience, the assumptions for multiple regression were not met. Therefore, a robust multiple regression was used.

The results from the robust multiple regression were non-significant, $F(12, 91) = 1.10$, $p > .05$. This shows that overall the predictor variables were not good at predicting ΔTrust (Table 3). The interactions between novices and outcome transparency was significant (t =

-2.00, $p < .05$). This shows that novices trust outcome transparency significantly less than procedural transparency.

Table 3.

Robust regression Analysis summary predicting the trust in DSS using transparency through explanations

Variable	Coef.	95% CI	Robust std. err.	t	p
Age	0.25	-0.06, 0.56	.16	1.62	.108
Sex					
Female	-0.01	-0.40, 0.38	.20	-0.04	.965
Prefer not to say	-0.46	-1.56, 0.63	.55	-0.84	.401
Experience with data from self trackers	0.01	-0.42, 0.44	.22	0.05	.961
Regression expertise	-0.02	-0.40, 0.36	.19	-0.11	.916
Work experience					
Novice	0.34	-0.30, 0.99	.33	1.05	.295
Experts	-0.81	-2.24, 0.63	.72	-1.12	.266
Transparency (0 = procedural, 1 = outcome)	0.29	-0.23, 0.81	.26	1.12	.265
Work experience * Transparency					
Novice	-1.08	-2.16, -0.01	.54	-2.00	.048*
Experts	0.09	-1.54, 1.73	.82	0.11	.911
Regression expertise * Transparency	0.12	-0.36, 0.59	.24	0.49	.629
Experience data * transparency	0.05	-0.59, 0.68	.32	0.15	.884
Constant	0.03	-0.37, 0.43	.20	0.15	.882

CI = Confidence interval, * = $p < .05$.

We did not find evidence for the hypotheses laymen trust procedural transparency more compared to outcome transparency (H2). Therefore, we could not reject the null-hypothesis laymen trust procedural transparency equally well as outcome transparency.

We did find a significant effect for novices and the type of transparency (Figure 10). Novices who saw procedural transparency trusted the recommendation of the DSS more than novices who saw outcome transparency compared to laymen (H3). Therefore, we reject the null-hypothesis novices trust procedural transparency equally well as outcome transparency.

We did not find evidence for the hypothesis experts trust outcome transparency more than procedural transparency (H4). We could thus not reject the null-hypothesis, experts trust procedural transparency equally well as outcome transparency.

Finally, no evidence was found to support the hypothesis that participants with more understanding of regression models trust outcome transparency more than procedural transparency (H5).

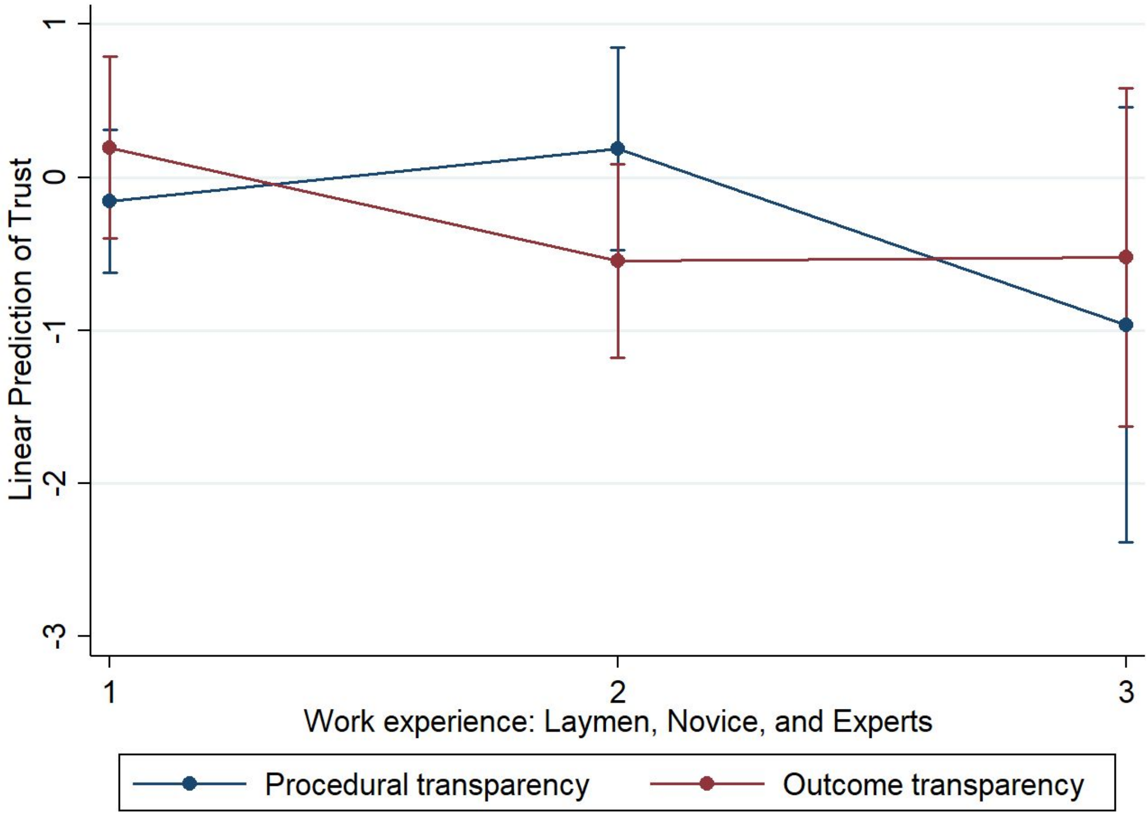

Figure 10. The Margins plot for the predicted values of standardized *Trust* for the three categories of *Work experience.*

The analysis of the data was also performed using work experience as a continuous variable, because the boundary between novice and expert is not well defined. The results from the robust multiple regression using work experience as a continuous variable were non-significant, $F(10, 93) = 0.91$, $p > .05$. Again showing that overall the predictor variables were not good at predicting ΔTrust. None of the variables or interactions were found to be significant ($p > .05$) (Table 4). The margins plot shows the expected direction for procedural transparency. However, not for outcome transparency (Figure 11). The difference between the lines was not statistically significant.

Table 4.
Robust regression Analysis summary predicting the Trust in DSS using Transparency through Explanations.

Variable	Coef.	95% CI	Robust std. err.	t	p
Age	0.24	-0.10, 0.57	.17	1.40	.166
Sex					
Female	-0.05	-0.47, 0.37	.21	-0.23	.820
Prefer not to say	-0.71	-1.98, 0.57	.64	-1.10	.273
Experience with data from self trackers	0.19	-0.22, 0.59	.21	0.90	.371
Regression expertise	-0.04	-0.46, 0.38	.21	-0.18	.856
Work experience	-0.33	-0.79, 0.14	.23	1.40	.165
Transparency (0 = procedural, 1 = outcome)	-0.04	-0.46, 0.38	.21	-0.18	.856
Work experience * Transparency	0.19	-0.38, 0.76	.29	0.67	.508
Regression expertise * Transparency	0.29	-0.18, 0.77	.24	1.23	.220
Experience data * transparency	0.29	-0.89, 0.30	.30	-1.00	.321
Constant	0.03	-0.38, 0.42	.20	0.12	.903

CI = Confidence interval, * = p < .05.

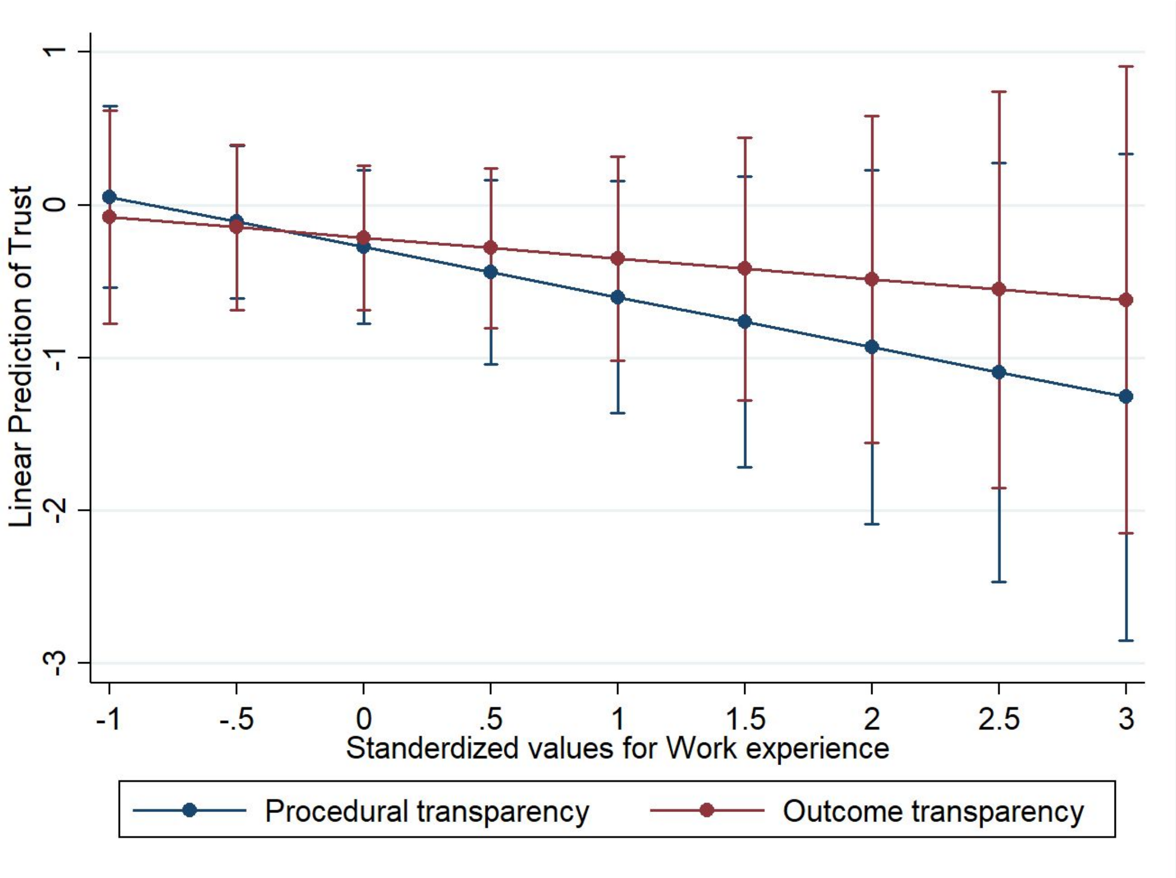

Figure 11. The Margins plot for the predicted values of standardized *Trust* for the standardized values of *Work experience.* Note that the range of *Work experience* indicates the unbalanced sample over *Work experience.*

When using work experience as a continuous variable, no evidence was found to reject the null-hypotheses.

Analysis of compliance

The effects of the interaction between work experience and the type of transparency on compliance with the recommendation of the participants were tested for all participants.

The assumptions for logistic regressions were met. There was no multicollinearity because there was no single variable with a variance inflation factor bigger than 5. The Pregibon's goodness-of-link test is based on the idea that the model is properly specified and that there is no link function that better predicts the outcome variable. Pregibon's goodness-of-link test was used to test the specificity of the model. The link test showed no problems with the interpretation of the logistic regression. Finally, the Hosmer–Lemeshow test for goodness of fit was performed and was found to be non-significant (Hosmer-Lemeshow chi2 (3) = 1.62, p = .66). Therefore, the assumption that the model fits everywhere equally well was met.

The logistic regression model was significant, the likelihood ratio Chi2 (9) = 23.77, p <.01. With a pseudo R squared of 18.3 %. To perform the likelihood ratio mode, only 102 participants were used because of some missing values in *Experience data*. Because *Experience data* (Z = 0.09, p = .93) and *Sex* (Z = -1.21, p = .22) were non-significant the model was run again without them so that the model included all participants. All assumptions were still met when performing the logistic regression without *Experience data* and *Sex*. The logistic regression correctly classified 77.48% of the observations.

The new likelihood ratio model was significant; the likelihood ratio Chi (6) = 25.26, $p < .001$ (Table 6). The interaction between *Work experience* and *Transparency* was also found to be a significant predictor for *Compliance* (OR = 2.68, 95% CI = 1.39, 5.16, Z = 2.94, $p < .05$), indicating that when participants had more work experience, they were more likely to comply with outcome transparency than procedural transparency (H4) and when participants had little work experience they were more likely to comply with procedural transparency than outcome transparency (H2 and H3). Therefore, we cannot reject the second, third and fourth hypotheses.

Also, Work experience was found to be a significant predictor of Compliance (OR = 0.39, 95% CI = 0.18, 0.83, Z = -2.45, $p < .05$), indicating Work experience had a negative effect on Compliance. No evidence was found for hypothesis five, therefore we cannot reject the null-hypothesis.

Table 6.
Logistic regression Analysis summary predicting the Compliance with the Recommendation of the DSS.

Variable	Odds ratio	95% CI	Std. err.	z	p
Age	1.03	0.48, 2.17	0.39	0.07	.945
Regression expertise	0.78	0.37, 1.65	0.30	-0.65	.515
Work experience	0.14	0.05, 0.45	0.08	-3.35	.001*
Transparency (0 = procedural, 1 = outcome)	1.66	0.65, 4.28	0.80	1.05	.295
Work experience * Transparency	7.17	1.93, 26.59	4.79	2.94	.003*
Regression expertise * Transparency	0.89	0.34, 2.32	0.44	-0.24	.813
Constant	1.56	0.75, 3.25	0.58	1.19	.233

CI = Confidence interval, * = $p < .05$.

Influential observations and outliers

Because of the skewness in the data, it was suspected that the model fits for the participants without any work experience as a health coach. Therefore, an extensive analysis of the influential observations and outliers was performed (Appendix C).

The analysis of the influential observations and outliers showed that in the regression model participants with more work experience as a health coach were seen as outliers with significant influence on the estimates for the coefficients.

Overall, the influential observations and outliers were older, had more work experience, and had less understanding of regression models. These traits were the traits that were expected from more experienced health coaches. Because the data collection is not distributed equally over work experience, the more experienced participants were seen as influential observations with a strong influence on the estimates of the parameters. Because this study was interested in the effects of work experience of health coaches and the interaction with the type of transparency, the following analysis will only use the participants with some experience working as a health coach (i.e., novice and expert health coaches).

Further analysis on participants with work experience as a health coach

The following analysis was performed on the subgroup of participants who have at least some work experience to see what the effects of the interaction between work experience and the type of transparency are. Thus, participants without any work experience as a health coach were excluded from the following analysis. The remaining sample consisted of 49 participants with a mean age of 26.1 with standard deviation 8.6, mean work experience of 3.1 years with standard deviation 2.8, out of which 22 participants saw procedural transparency and 27 saw outcome transparency.

The model assumptions for multiple regression were accepted. The assumptions for normally distributed residuals was tested using the Shapiro-Wilk W test for normal data, and the Skewness/Kurtosis tests for Normality. Both tests did not reject the null hypothesis ($W = 0.97$, $p = 0.27$; $Chi2 = 0.71$, $p = .70$) and therefore normality of the residuals was assumed. Cameron and Trivedi's decomposition of IM-test and Breusch-Pagan / Cook-Weisberg test for heteroskedasticity were used to test the assumption of homoscedasticity. Both test could not reject the null-hypotheses (Cameron and Trivedi $Chi2 = 55.40$, $p = .22$; Breusch-Pagan $Chi(1) = 3.98$, $p = .05$) and therefore homoscedasticity was assumed. Finally, the assumption of multicollinearity was checked by calculating the variance inflation factor. No single predictor variable had a variance inflation factor bigger than five, and the mean-variance inflation factor was lower than two. Therefore, the assumption of multicollinearity was met. Thus, the following multiple regression was performed (Table 7).

Table 7.
Regression Analysis summary predicting the Trust in CDSS using Transparency through Explanations for participants with Work Experience as a Health Coach.

Variable	Coef.	95% CI	Robust std. err.	t	p
Age	0.28	-0.13, 0.70	.20	1.39	.175
Sex					
Female	-0.31	-0.52, 1.15	.41	0.77	.448
Prefer not to say	-0.70	-3.24, 1.84	1.25	-0.56	.577
Experience with data from self trackers	0.44	-0.65, 1.53	.53	0.82	.416
Regression expertise	-0.04	-0.46, 0.53	.24	-0.15	.882
Work experience	-1.12	-1.92, -0.31	.40	-2.82	.008*
Transparency (0 = procedural, 1 = outcome)	-1.63	-2.84, -0.43	.40	-2.76	.010*
Work experience * Transparency	1.27	0.28, 2.26	.46	-2.76	.014*
Regression expertise * Transparency	0.11	-0.53, 0.75	.31	0.34	.734
Experience data * transparency	-0.16	-1.57, 1.26	.70	-0.23	.822
Constant	0.82	-0.16, 1.80	.48	1.71	.097

CI = Confidence interval, * = $p < .05$.

The results for the multiple regression analysis were non-significant, $F(10, 32) = 1.61$, $p = .15$. Showing that overall the predictor variables were not good at predicting ΔTrust. The results for the interaction term between *Work experience* and *Transparency* was significant ($t = 2.61$, $p < .05$), indicating that when participants had more work experience, they trusted outcome transparency more compared to procedural transparency (H4) and when participants had little work experience they trusted procedural transparency more than outcome transparency (H3). Therefore, when we only use participants with at least some work experience, we cannot reject the third and fourth hypotheses. The marginal plot for this interaction shows the relationship between work experience and the type of transparency (*Figure 12)*.

The results for the *Work experience* was significant in the multiple regression model ($t = -2.82$, $p < .05$), indicating a significant difference between the trust in the recommendation and the work experience of the participant. When a participant had more work experience, they trusted the recommendation less. The results for *Transparency* were significant (t = -2.76, p < .05). This significant variable indicated that the participants trusted procedural outcome transparency more.

No evidence for the fifth hypothesis could be found. Therefore, the null-hypothesis, participants with more knowledge about regression and formulas trust procedural and outcome transparency equally well, could not be rejected.

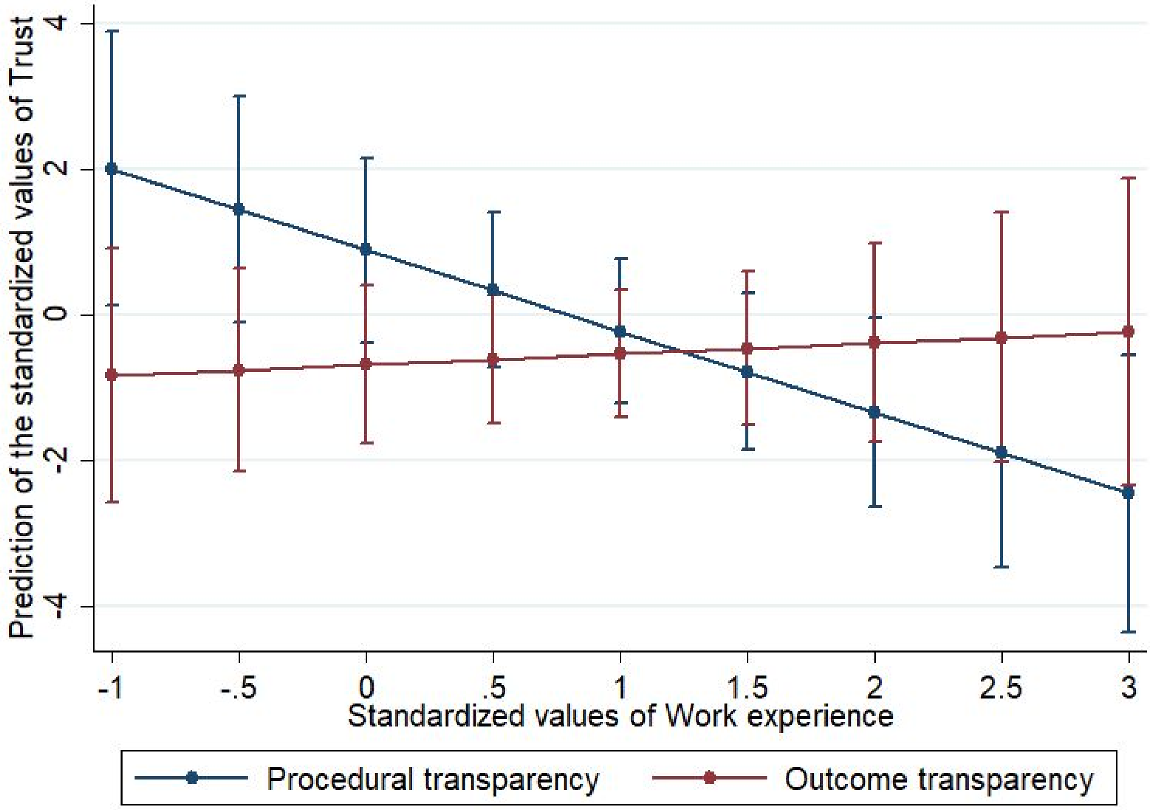

Figure 12. The Marginsplot for the Multiple Regression on Trust for the Participants with Work Experience.

The effects of the interaction between work experience and the type of transparency on compliance with the recommendation of the participants were tested for the subgroup of participants who have at least some work experience as a health coach.

The assumptions for logistic regressions were met. There was no multicollinearity because there was no single variable with a variance inflation factor bigger than 5. The Pregibon's goodness-of-link test is based on the idea that the model is properly specified and that there is no link function that better predicts the outcome variable. Pregibon's goodness-of-link test was used to test the specificity of the model. The link test showed no problems with the interpretation of the logistic regression. Finally, the Hosmer–Lemeshow test for goodness of fit was performed and was found to be non-significant (Hosmer-Lemeshow chi2 (38) = 26.47, p = .16). Therefore, the assumption that the model fits everywhere equally well was met.

The logistic regression model was significant, the likelihood ratio Chi2 (9) = 16.66, p <.05. With a pseudo R squared of 24.5 % (Table 8). The logistic regression correctly classified 79.59% of the observations.

The interaction between *Work experience* and *Transparency* was also found to be a significant predictor for *Compliance* (OR = 35.99, 95% CI = 1.74, 745.69, Z = 2.32, $p < .05$), indicating that when participants had more work experience, they were more likely to comply with outcome transparency than procedural transparency (H4) and when participants had little work experience they were more likely to comply with procedural transparency than outcome transparency (H2 and H3). Therefore, we cannot reject the second, third and fourth hypotheses. No evidence was found for hypothesis five, therefore we cannot reject the null-hypothesis..

Table 8.
Logistic regression Analysis summary predicting the Compliance with the Recommendation of the DSS for the subgroup of participants who have at least some Work Experience as a Health Coach.

Variable	Odds ratio	95% CI	Std. err.	z	p
Age	0.66	0.23, 1.85	0.35	-0.79	.427
Regression expertise	0.90	0.31, 2.52	0.47	-0.22	.826
Work experience	0.13	0.01, 1.92	0.19	-1.47	.141
Transparency (0 = procedural, 1 = outcome)	0.37	0.15, 3.11	0.41	-0.90	.367
Work experience * Transparency	35.99	1.74, 745.70	55.66	2.32	.020*
Regression expertise * Transparency	0.60	0.15, 2.37	0.42	-0.72	.469
Constant	1.67	0.30, 9.25	1.46	0.59	.555

CI = Confidence interval, * = $p < .05$.

Qualitative results

In the following section, the open questions are discussed. In the first question, participants were asked what information they were missing needed to give good coaching advice. In the second question, participants were asked for any comments or remarks about the tool and the experiment. The answers to the open questions were labeled with codes and the frequency of these codes was recorded to discover if the participants had enough information to formulate an advice. Also to find if the missing information could be provided with PGD or in a conversation. Moreover, the quotes were qualitatively examined to discover the opinions of the participants.

The first open question: missing information

The participants saw the first open question after using the tool with low transparency. The goal of the question was to discover if participants were still missing any data that they needed in order to formulate good coaching advice. The following question was asked:

'Do you have enough information from the tool to give Sam a good coachings advice? If not, can you indicate which information is missing?'

The qualitative data on what information the participants were missing from the tool was examined. Several codes were identified (e.g. stress and mental health, sleep times, and clients subjective sleep quality). These codes were grouped together in themes (environment, sleep, and exercise). These themes represented the type of information that the participants were missing.

Many of the participants noted that they missed information about stress and mental health before they were able to give good coaching advice. For example, participant 53 mentioned: *"There is no information about Sam's feelings. For instance, bad sleep may be due to stress."* and participant 62 noted: *"There might be other issues such as stress from his work/study/relation that could have a negative influence on his sleep. This information is crucial in giving solid advice, but is harder to gather in a number.".*

Participants also missed more detailed information about the sleep times from the client: "*[it] would be useful to see what time exactly he [the client] sleeps*", participant 63 noted. Moreover, they also missed information on the times the client drank coffee and alcohol: "*I feel like the information that is missing would be the time of day he (the client) drinks coffee or alcohol*", participant 73 remarked.

The participants also missed more personal information about the client, for example, information on the client's personal goals: "*It is unknown what his personal goals are, what is his request for help, and what is his problem. So, it is not possible to give targeted advice.*", participant 2 said. Also some participants noted that they wanted to know in what kind of environment the client lived: "*What is the social context and environment in which Sam lives*", participant 111 asked.

Some of the missing information could be solved with more data. Furthermore, participants also noted that the data was a good starting point for a conversation. However, the data alone was not enough and that the data was limiting their thinking: "*Looking at the available data, exercise seems to provide the participant with the most sleep because he sleeps best after he exercised. However, I notice that the dashboard is limiting my thinking over possible causes to only the available information while there could be much more going on: stress (young children?) or trouble at work*", participant 8 noted.

On the one hand, themes that were more easily tracked as PGD were identified. This included for example sleep, exercise, and food and drinks. On the other hand, there were themes identified that were tracked more easily by a conversation between coach and client including stress and mental health and environment. *Figure 13* demonstrates all the discovered codes, how they are grouped into themes, and whether the missing information could be best tracked with data or a conversation.

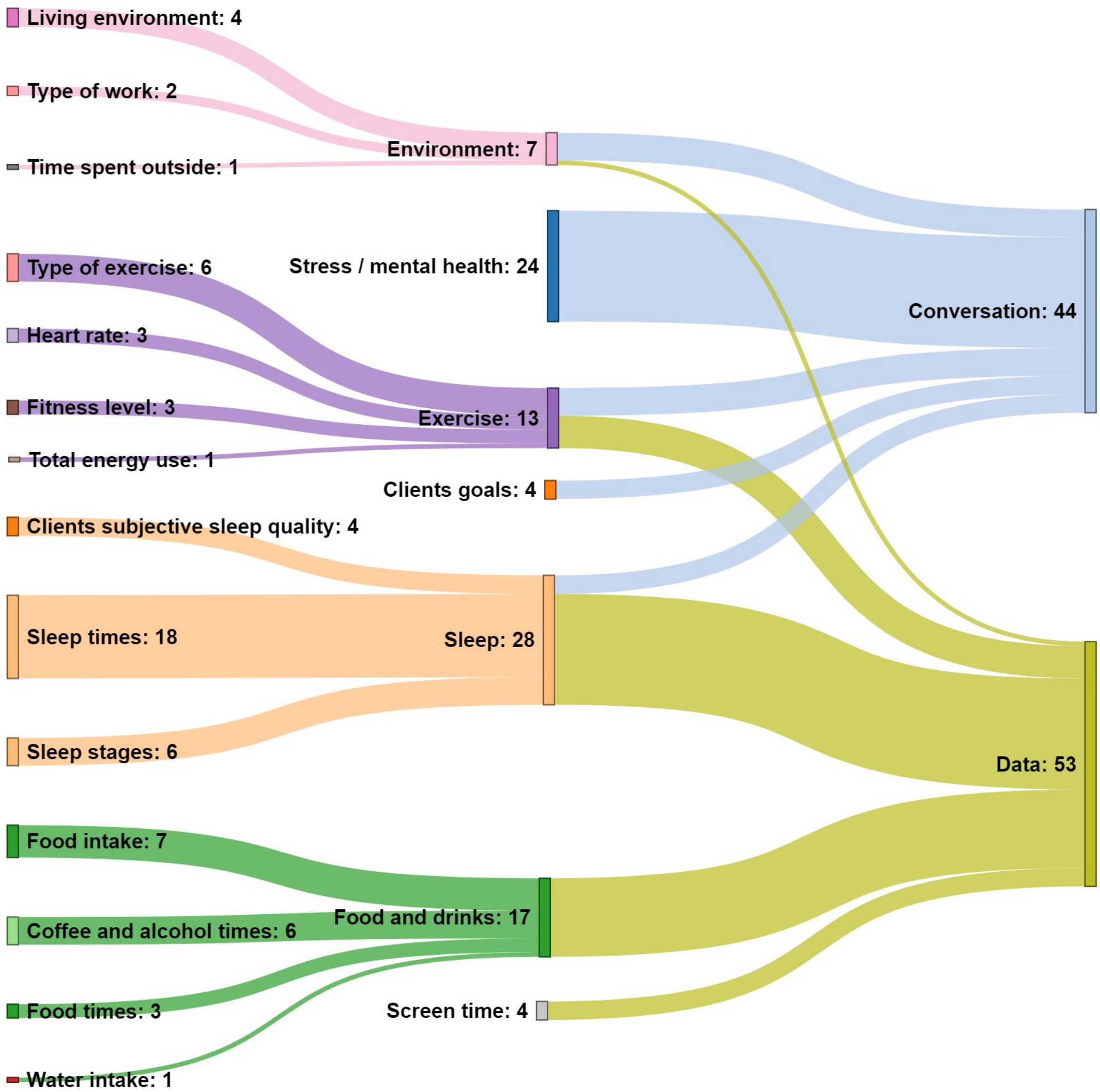

Figure 13. The Codes and Themes from the open question. *"Do you have enough information from the tool to give Sam a good coachings advice? If not, can you indicate which information is missing?"*.

From the 97 observations that were made in the data, 44 were best solved with a conversation and 53 with more data. The most occurring code that was discovered was Stress / mental health (24). These observations were mostly personal questions that the participants wanted to ask the client. This showed that some of the missing information could have been provided by the dashboard.

Overall, the first question showed that the participants needed more information in order to give good coaching advice. The three most prominent information themes they were missing were stress and mental health information, sleep information, and food and drinks information. The stress and mental health theme was entirely missing from the tool.

Sleep and food and drinks information were available in the tool. However, the participants wanted more detailed information in order to give good coaching advice. From the missing data, some could have been measured with a self-tracker and some would be best gathered in a face-to-face conversation. The data that could have been tracked with self-trackers could have been shown on the dashboard.

The second open question: comments about the tool

The participants were shown the second open question after filling in the demographic questions. This question was asked to discover whether the participants had any remarks about the tool. The following question was asked:

"Do you have any questions or comments about the study, or about the tool?"

Some interesting comments were chosen to be discussed in the following paragraph. Both positive and negative comments about the tool are highlighted. Overall, the second question showed that the participants did appreciate the tool and the recommendation. However, there were also concerns about the accuracy of the data and the system. Furthermore, some participants mentioned that they were not able to give advice solely based on the tool.

Participant 1 remarked that the extra information in the procedural transparency fostered more trust by opening the black-box: "*The extra information from the tool increases the trust because it no longer is a black-box model* ". Participant 10 said it is not possible to give advice solely based on data from wearables, but it could be a tool to help give the client insights: "*Advice solely based on data from the tool should never be given by any health coach. Which does not take away that self-monitoring or self-measurement is desirable and that it can be of great support during the recovery* " (procedural transparency).

Many participants were positive about the tool. For example, participant 12 said: "*This tool is great, I would definitely work with it if it were easily accessible*." (procedural transparency). However, many participants also said the tool would not be enough to give good coaching advice and more personal information is needed: "*I would take the recommendation from the tool into account in my final advice, but it would not be the most important parameter. More important is to know the personal situation and from there give advice to the client that is achievable*", participant 13 remarked (procedural transparency). Other participants questioned the accuracy of the data: "*How reliable is a smartwatch to accurately measure sleep?*", participant 16 asked (outcome transparency); "*How is it possible that an activity tracker can measure the time the client slept? Is it the actual sleep time, and how reliable is it?*", participant 27 noted (outcome transparency). Other participants questioned whether it was possible to give good coaching advice based solely on the tool: "*I think it is not possible to give advice solely based on the tool.*", participant 29 said (outcome transparency). Finally, some participants did not see the difference between the conditions: "*The explanation of the recommendation could maybe more clear. It was not immediately clear to me what the difference was between the two tools*", participant 49 noted (procedural outcome).

Discussion

The goal of this study is to investigate the effects of transparency in the recommendations of a decision support system (DSS) that uses patient-generated data (PGD). The findings aim to bring DSS using PGD a step closer to being adopted by health coaches.

Prior research mainly focuses on the opportunities of PGD within the clinical healthcare setting. These technologies also offer great opportunities for health coaches, but research remains limited. Therefore, this study aims to contribute to the lack of empirical research by examining the relationship between different types of transparency in DSS and trust by health coaches within these systems. Also, this study looked at the relationship between the kind of transparency and work experience of the health coaches. Finally, this study looked at how knowledge about regressions and formulas influenced the trust DSS when different types of transparency were provided.

Finding of the current study

Quantitative findings

In general, this study did not find support for the first hypothesis; transparency through explanation increases trust in a DSS that used PGD. Therefore, we cannot accept the first hypothesis. There was no statistical difference in trust between low transparency and transparency through explanations.

These results are not in line with previous findings (Beck et al., 2007; Dzindolet et al., 2002; Kilzelcec, 2016, Lyons et al., 2017; Nilashi et al., 2016), where transparency through explanations increased trust in automated systems. Similar operationalizations of transparency were used by Kilzelcec (2016), Lyons et al. (2017), and Nilashi et al., (2016). In these studies, they did find effects for the type of explanation that provided transparency and trust in automated systems. Therefore, the explanation for the missing effect may be found elsewhere.

An explanation for the absence of an effect could be the difference between the trust in the low transparency condition and the explanation condition. In the low transparency condition, only visualizations were provided as transparency. The condition that used explanations provided either procedural or outcome explanations to increase the transparency. The difference in trust between these conditions was small. Because visualization is already a source of transparency, it may already provide enough trust (Fekete et al., 2008; Saket et al., 2018).

This study also looked at the interaction between work experience and the type of explanation that provide transparency. It is known that expert users require different forms of transparency than laymen and novices (Arnold et al., 2006; Cheng et al., 2019; Desai et al., 2019; Schaffer et al., 2019). First, it was expected that both the layman and the novice health coaches would trust explanations regarding the procedural steps more than explanations regarding the outcome. Also, it was expected that experts, participants with work experience as a health coach, would trust the recommendation of the DSS more when an explanation regarding the outcome was provided as opposed to an explanation regarding the procedural steps. Finally, it was expected that participants with more knowledge about regressions and formulas would trust explanations regarding the outcome more compared to explanations regarding the procedural steps.

Evidence was found for the hypothesis that novice health coaches trust procedural transparency more than outcome transparency (H3). This result was in line with the findings from Kilzelcec (2016), where procedural transparency was also more trusted than outcome transparency for novice users. However, when work experience was used as a continuous variable in the regression model, the effect disappeared. No evidence was found to support the other hypotheses (H2, H4, and H5).

An analysis of compliance was performed to test the hypotheses as compliance or the willingness to depend is often used as a measure for trust. The analysis tested if the participants agreed with the recommendation of the system. The interaction between *Work experience* and *Transparency* was found to be a significant predictor for *Compliance*. This indicated that when participants had more work experience, they were more likely to comply with outcome transparency than procedural transparency (H4). Moreover, when participants had little work experience they were more likely to comply with procedural transparency than outcome transparency (H2 and H3). Therefore, the second, third and fourth hypotheses could not be rejected.

This is in line with the findings of previous research where novice users of automated systems trusted procedural transparency more than outcome transparency and expert users trusted outcome transparency more than procedural transparency (Arnold et al., 2006; Cheng et al., 2019; Desai et al., 2019; Schaffer et al., 2019). This shows that experienced health coaches trust recommendations from DSS that use outcome transparency more than the recommendations from DSS that use procedural transparency. Also, less experienced health coaches trust recommendations from DSS that use procedural transparency more than recommendations from DSS that use outcome transparency.

The skewness of the sample was expected to be an explanation for the missing effects on trust. It was expected that the model fitted for the participants without any working experience. Also, because previous studies specifically used novices and experts without laymen to test the effects of transparency in automated systems, the analysis was performed again (Arnold et al., 2006; Cheng et al., 2019; Desai et al., 2019; Schaffer et al., 2019). This time only participants with at least some working experience as a health coach were included (i.e., only novices and experts) to test the third and fourth hypotheses.

When only the novices and experts health coaches were used ($n = 49$), the third hypothesis and fourth hypothesis could not be rejected. This shows that novice health coaches trusted procedural transparency more than outcome transparency and that expert health coaches trusted outcome transparency more than procedural transparency.

These findings are again in line with previous research (Arnold et al., 2006; Cheng et al., 2019; Desai et al., 2019; Schaffer et al., 2019), where more experienced users trusted transparency in recommendations more when outcome transparency was provided. Also, novice users trusted recommendations more when procedural transparency was provided.

Furthermore, work experience was a significant predictor of trust. Participants with more work experience had less trust in the system. This result is in line with previous findings where the trust of experts is lower than the trust of novices in an automated system (Hoff and Bashir, 2014). Also, the type of transparency that was used was a significant predictor of trust. The participants trusted procedural transparency more than outcome transparency.

Finally, the analysis was performed on compliance for the participants with at least some work experience as a health coach. The interaction between Work experience and Transparency was again found to be a significant predictor for Compliance. This indicates that when participants had more work experience, they were more likely to comply with outcome transparency than procedural transparency (H4). Moreover, when participants had little work experience they were more likely to comply with procedural transparency than outcome transparency (H2 and H3). Therefore, the second, third and fourth hypotheses could not be rejected.

This study could not find any evidence for the effects of knowledge about regression models and formulas and the type of transparency (H5). The lack of these findings might be explained by the overall knowledge of the participant with regards to regression models and formulas. Most participants were not familiar or only somewhat familiar with regression models. Also, the knowledge about regression models and formulas may not have been enough for the participants to not get confused. Therefore, similar effects as the effects found by Kizilcec (2016) may have occurred where the confusion distracted participants from the explanation.

Qualitative findings

Two open questions were analyzed in order to further understand the needs of the participants regarding the DSS and to discover how future research could improve on this experiment.

Generally, the first question showed that the participants want more information in order to give good coaching advice. From the missing data, about half could have been measured with a self-tracker, and the other half would best be gathered in a conversation. These findings seem to indicate that there is some information the participants need that was not available as PGD. Also, it shows that the participants need more information in general in order to give good coaching advice. More personal information was mentioned such as, personal goals and mental wellbeing.

This hints that data alone is not enough to give coaching advice and that a coach-client relationship remains essential. It shows that DSS using PGD cannot replace coach-client session but rather complement it.

The second open question showed that the participants did appreciate the tool and its advice. Some participants mentioned the transparency opened the “black box,” and they trusted the tool more. However, the reliability of the data was questioned together with the accuracy of the system. This doubt shows that transparency is needed, not only in the interpretation of data, but also in data acquisition and performance of the system.

Limitations and recommendations for future research

To more reliably understand the needs of health coaches, future research should aim to recruit more experienced health coaches. The current sample underrepresented individuals with significant work experience as a health coach. The sample was therefore unbalanced and the findings of the effect of interaction between *Transparency* and *Work experience* less robust.

Additionally, future research could design a client case where the recommendation is not too noticeable. In this experiment, the recommendation showed that exercise would be the best first step in helping the client. The overall trust the participants had in the recommendation of the DSS was high. Before the participants saw any transparency through explanations, already 49 out of 111 participants complied with the recommendation. Kilzelcec 2016, showed that when the expectations are met, the trust in automation already is high and that procedural and outcome transparency do not increase trust. In this study, no effect was found for the increase in trust through explanations (H1). To better understand the effects of explanations, either procedural or outcome explanations, a case needs to be designed where the recommendation can be more difficult to identify in the data. This study proposes that future research designs a case where the solution is an interaction of two or more factors. An interaction between factors would make it more difficult to confirm the recommendation by looking at the dashboard. Procedural and outcome transparency could potentially explain this interaction and thus have a greater effect on trust.

A more complicated case is a reasonable way of testing procedural and outcome transparency because the main interest is in situations where the solution is not easy. Only the cases where there is a disagreement between the health coach and the DSS can transparency through explanations increase trust.

Conclusion

This study set out to understand the transparency needs of health coaches in decision support systems (DSS) that use patient-generated data (PGD). The primary contribution of this study is to improve understanding how work experience and the type of transparency through explanations interact to affect the trust health coaches have in recommendations from DSS.

On the one hand, this study shows that more experienced health coaches trust recommendations from DSS more when explanations about the outcome are provided compared to explanations about the procedure. On the other hand, less experienced health coaches trust recommendations from DSS more when procedural explanations are provided compared to outcome explanations.

Although this study found effects for the interaction between the work experience of health coaches and the type of explanations that were used to provide transparency, more research is needed. Future research could benefit from exploring situations where the expectation of participants is not met in order to better understand the effects of the type of explanation that is used to provide transparency.

While this study clearly supported the current state of the art in this field, it also provided new insights and a new perspective by focussing on health coaches. Hence, these results provide developers of DSS with insights on how to effectively design DSS with regards to transparency for health coaches with different levels of work experience, and thus facilitate its implementation and use of the benefits provided.

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access, 6, 52138–52160.* https://doi.org/10.1109/ACCESS.2018.2870052

Alsos, O. A., Das, A., & Svanæs, D. (2012). Mobile health IT: The effect of user interface and form factor on doctor-patient communication. *International Journal of Medical Informatics, 81(1), 12–28.* https://doi.org/10.1016/j.ijmedinf.2011.09.004

Arnold, V., Clark, N., Collier, P. A., Leech, S. A., & Sutton, S. G. (2006). The differential use and effect of knowledge-based system explanations in novice and expert judgment decisions. *Mis Quarterly, 79-97*.

Beck, H. P., Dzindolet, M. T., & Pierce, L. G. (2007). Automation Usage Decisions: Controlling Intent and Appraisal Errors in a Target Detection Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49(3), 429–437.* https://doi.org/10.1518/001872007X200076

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Detecting and assessing collinearity. *Regression diagnostics: Identifying influential data and sources of collinearity, 85-191.*

Brusco, N. K., & Watts, J. J. (2015). Empirical evidence of recall bias for primary health care visits. *BMC Health Services Research, 15(1).* https://doi.org/10.1186/s12913-015-1039-1

Cardinali, C. (2013). Observation influence diagnostic of a data assimilation system. In Data Assimilation for Atmospheric, *Oceanic and Hydrologic Applications (Vol. II) (pp. 89-110)*. Springer, Berlin, Heidelberg.

Cecaj, A. (2016, February 13). Data sets of any type: some links. - Algorithms & Data Fusion - Quora. Retrieved March 28, 2019, from https://www.quora.com/q/ktgvobsmnbchdxii/Data-sets-of-any-type-some-links

Charleer, S., Gutiérrez Hernández, F., & Verbert, K. (2018). Supporting job mediator and job seeker through an actionable dashboard. *In Proceedings of the 24th IUI Conference on Intelligent User Interfaces. ACM.*

Cheng, H.-F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019). Explaining Decision-Making Algorithms through UI. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19.* https://doi.org/10.1145/3290605.3300789

Chung, C. F., Dew, K., Cole, A., Zia, J., Fogarty, J., Kientz, J. A., & Munson, S. A. (2016, February). Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data. *In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (pp. 770-786). ACM.*

Chung, C., Cook, J., Bales, E., Zia, J., & Munson, S. A. (2015). More Than Telemonitoring: Health Provider Use and Nonuse of Life-Log Data in Irritable Bowel Syndrome and Weight Management. *Journal of Medical Internet Research, 17(8).* https://doi.org/10.2196/jmir.4364

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., . . . Shekelle, P. G. (2006). Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. *Annals of Internal Medicine, 144(10), 742.* https://doi.org/10.7326/0003-4819-144-10-200605160-00125

Claes, S., & Moere, A. V. (2015, June). The role of tangible interaction in exploring information on public visualization displays. *In Proceedings of the 4th International Symposium on Pervasive Displays (pp. 201-207). ACM.*

Degtiarev, K. Y., & Remnev, N. V. (2016). Linguistic Resumes in Software Engineering: The Case of Trend Summarization in Mobile Crash Reporting Systems. *Procedia Computer Science, 102, 121–128*. https://doi.org/10.1016/j.procs.2016.09.378

Deep Sleep: How to Get More of it – *American Sleep Association.* (2019, May 5). Retrieved May 10, 2019, from https://www.sleepassociation.org/about-sleep/stages-of-sleep/deep-sleep/

Dejonghe, L. A. L., Becker, J., Froboese, I., & Schaller, A. (2017). Long-term effectiveness of health coaching in rehabilitation and prevention: A systematic review. *Patient Education and Counseling*, 100(9), 1643–1653. https://doi.org/10.1016/j.pec.2017.04.012

Deokar, A. V., & Sarnikar, S. (2014). Understanding process change management in electronic health record implementations. *Information Systems and e-Business Management*, 14(4), 733–766. https://doi.org/10.1007/s10257-014-0250-7

Desai, P. M., Mitchell, E. G., Hwang, M. L., Levine, M. E., Albers, D. J., & Mamykina, L. (2019). Personal Health Oracle. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19.* https://doi.org/10.1145/3290605.3300600

Dzindolet, M., Pierce, L., Peterson, S., Purcell, L., Beck, H., & Beck, H. (2002, September). The influence of feedback on automation use, misuse, and disuse. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 46, No. 3, pp. 551-555)*. Sage CA: Los Angeles, CA: SAGE Publications.

Eslami, M., Krishna Kumaran, S. R., Sandvig, C., & Karahalios, K. (2018, April). Communicating algorithmic process in online behavioral advertising. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 432). ACM.*

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39(11), 27-34.*

Fekete, J., Van Wijk, J. J., Stasko, J. T., & North, C. (2008). The Value of Information Visualization. *Lecture Notes in Computer Science, 1–18.* https://doi.org/10.1007/978-3-540-70956-5_1

Gagnon, M., Ngangue, P., Payne-Gagnon, J., & Desmartis, M. (2015). m-Health adoption by healthcare professionals: a systematic review. J*ournal of the American Medical Informatics Association, 23(1), 212–220.* https://doi.org/10.1093/jamia/ocv052

Gray, J. P., Arlinghaus, K. R., & Johnston, C. A. (2018). The Potential Contribution of Current Health Care Zeitgeists to the Success of Health and Wellness Coaches. *American Journal of Lifestyle Medicine*, 12(6), 456–458. https://doi.org/10.1177/1559827618792496

He, C., Parra, D., & Verbert, K. (2016). Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications, 56, 9-27.* https://doi.org/10.1016/j.eswa.2016.02.013

Hoff, K. A., & Bashir, M. (2014). Trust in Automation. Human Factors: *The Journal of the Human Factors and Ergonomics Society, 57(3), 407–434.* https://doi.org/10.1177/0018720814547570

Hong, M. K., Lakshmi, U., Olson, T. A., & Wilcox, L. (2018). Visual ODLs. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18.* https://doi.org/10.1145/3173574.3174050

Huba, N., & Zhang, Y. (2012). Designing Patient-Centered Personal Health Records (PHRs): Health Care Professionals' Perspective on Patient-Generated Data. *Journal of Medical Systems, 36(6), 3893–3905.* https://doi.org/10.1007/s10916-012-9861-z

Kacprzyk J., Zadrożny S. (2009) Supporting Decision Making via Verbalization of Data Analysis Results Using Linguistic Data Summaries. In: Rakus-Andersson E., Yager R.R., Ichalkaranje N., Jain L.C. (eds) Recent Advances in Decision Making. *Studies in*

Computational Intelligence, vol 222. Springer, Berlin, Heidelberg

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science, 4(6), 401–405*. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Kellermann, A. L., & Jones, S. S. (2013). What It Will Take To Achieve The As-Yet-Unfulfilled Promises Of Health Information Technology. *Health Affairs, 32(1), 63 - 68.* https://doi.org/10.1377/hlthaff.2012.0693

Kizilcec, R. F. (2016, May). How much information?: Effects of transparency on trust in an algorithmic interface. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 2390-2395). ACM.* http://dx.doi.org/10.1145/2858036.2858402

Krummel, T. M. (2019). The Rise of Wearable Technology in Health Care. *JAMA Network Open, 2(2).* https://doi.org/10.1001/jamanetworkopen.2018.7672

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2006). Is It Time for Clinicians to Routinely Track Patient Outcome? A Meta-Analysis. *Clinical Psychology: Science and Practice, 10(3), 288–301.* https://doi.org/10.1093/clipsy.bpg025

Lind, E. A., & Tyler, T. R. (1988). The social psychology of procedural justice. *Springer Science & Business Media*.

Lyons, J. B., Sadler, G. G., Koltai, K., Battiste, H., Ho, N. T., Hoffmann, L. C., ... & Shively, R. (2017). Shaping trust through transparent design: theoretical and experimental guidelines. *In Advances in Human Factors in Robots and Unmanned Systems (pp. 127-136). Springer, Cham.*

Matt, G. E., Vázquez, C., & Campbell, W. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review, 12(2), 227–255.* https://doi.org/10.1016/0272-7358(92)90116-p

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, 20(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*, 13(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Neff, G., & Nafus, D. (2016). Self-Tracking (Rev. ed.). *Cambridge, USA: MIT Press.*

Nilashi, M., Jannach, D., bin Ibrahim, O., Esfahani, M. D., & Ahmadi, H. (2016). Recommendation quality, transparency, and website quality for trust-building in recommendation agents. *Electronic Commerce Research and Applications, 19, 70-84.*

Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction, 27(3–5), 393–444*.
https://doi.org/10.1007/s11257-017-9195-0

O'Broin, A., & Palmer, S. (2006). The coach-client relationship and contributions made by the coach in improving coaching outcome. *The Coaching Psychologist, 2(2), 16-20.*

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 30(3), 286–297*.
https://doi.org/10.1109/3468.844354

Parra, D., & Brusilovsky, P. (2015). User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies, 78, 43–67.*
https://doi.org/10.1016/j.ijhcs.2015.01.007

Raj, S., Lee, J. M., Garrity, A., & Newman, M. W. (2019). Clinical Data in Context: Towards Sensemaking Tools for Interpreting Personal Health Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(1), 22.*

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not So Different After All: A Cross-Discipline View Of Trust. *Academy of Management Review*, 23(3), 393–404.
https://doi.org/10.5465/amr.1998.926617

Rutjes, H., Willemsen, M. C., Van Kollenburg, J., Bogers, S., & IJsselsteijn, W. A. (2017, May). Benefits and costs of patient-generated data, from the clinician's and patient's perspective. *In Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare (pp. 436-439). ACM.*

Rutjes H., Willemsen, M. C., and IJsselsteijn, W. A. (2019a). Beyond Behavior: The Coach's Perspective on Technology in Health Coaching. *In 2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK. ACM, NY, NY, USA. Paper 670, 14 pages.*
https://doi.org/10.1145/3290605.3300900

Rutjes, H., Willemsen, M. C., & IJselsteijn, W. A. (2019b). Considerations on Explainable AI and Users' Mental Models. *In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk, ACM.*

Saket, B., Moritz, D., Lin, H., Dibia, V., Demiralp, C., & Heer, J. (2018). Beyond heuristics: Learning visualization design. *arXiv preprint arXiv:1807.06641.*

Schaefer, Kristin E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58(3), 377–400. https://doi.org/10.1177/0018720816634228

Schaffer, J., Playa Vista, C. A., O'Donovan, J., Michaelis, J., Adelphi, M. D., Raglin, A., & Höllerer, T. (2019, March). I can do better than your AI: expertise and explanations. *In Proceedings of the 24th International Conference on Intelligent User Interfaces (pp. 240-251). ACM.*
https://doi.org/10.1145/3301275.3302308

Shapiro, M., Johnston, D., Wald, J., & Mon, D. (2012). Patient-generated health data. RTI International, April.

Sittig, D. F., Wright, A., Osheroff, J. A., Middleton, B., Teich, J. M., Ash, J. S., . . . Bates, D. W. (2008). *Grand challenges in clinical decision support. Journal of Biomedical Informatics, 41(2), 387–392.*
https://doi.org/10.1016/j.jbi.2007.09.003

Slaapproblemen | Thuisarts. (n.d.). Retrieved May 10, 2019, from https://www.thuisarts.nl/slaapproblemen

Stewart, M. A. (1995). Effective physician-patient communication and health outcomes: a review. CMAJ: Canadian Medical Association Journal, 152(9), 1423.

Thudt, A., Hinrichs, U., & Carpendale, S. (2012, May). The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1461-1470). ACM.*

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems, 23(4), 217-246.*

West, P., Van Kleek, M., Giordano, R., Weal, M. J., & Shadbolt, N. (2018). Common Barriers to the Use of Patient-Generated Data Across Clinical Settings. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18,* https://doi.org/10.1145/3173574.3174058

Van Dijk, E. K., IJsselsteijn, W., & Westerink, J. (2016). Deceptive visualizations and user bias. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16.*
https://doi.org/10.1145/2968219.2968326

Zhu, H., Colgan, J., Reddy, M., & Choe, E. K. (2017). Sharing Patient-Generated Data in Clinical Practices: An Interview Study. *AMIA ... Annual Symposium proceedings. AMIA Symposium, 2016, 1303-1312.*

Appendix

Appendix A, Questions in the survey

Table 9.
The questions with the possible answer options.

Variable	Question	Answer options
Recommendation choice	What would you focus on most in your advice to Sam? In other words, which advice do you think would benefit Sam the most?	Advice regarding: Sleep / Coffee / Alcohol / Light move minutes / Exercise / Inactive minutes / Steps
Advice	What would be your advice regarding this subject?	[Open question]
Need more information	Do you have enough information from the tool to give Sam a good coachings advice? If not, can you indicate which information is missing?	[Open question]
Trust1	I feel comfortable using this tool	Totally disagree - Totally agree (7 point Likert)
Trust2	I believe that these types of tools are effective	Totally disagree - Totally agree (7 point Likert)
Trust3	I think using the tool's recommendation leads to a positive outcome	Totally disagree - Totally agree (7 point Likert)
Trust4	I do not trust this recommendation	Totally disagree - Totally agree (7 point Likert)
Trust5	I believe this recommendation is useful	Totally disagree - Totally agree (7 point Likert)
Trust6	If I were a health coach, I would not intend to use similar tools in the future	Totally disagree - Totally agree (7 point Likert)
Age	What is your age?	[Number between 1 and 99]
Sex	What is your gender?	Man / Women / Prefer not to say
Work experience	How much experience do	No experience / Less than 1

	you have working with individual clients?	year / Between 1 and 3 years / Between 3 and 5 years / Between 5 and 10 years / More than 10 years
Same client	Do you coach clients with problems similar to Sam? (That is, sleep or stress-related issues.)	Yes, often / Yes, sometimes / No, never
Work with data	Do you use data from self-trackers/wearables when coaching?	Yes, I use this on my own initiative / Yes, but only when the client initiates it / No
Experience data	How much experience do you have with viewing data from self-trackers/wearables? For example, because you or someone you know is wearing one.	A lot of experience, I have often viewed these data / Little experience, I have sometimes viewed these data / No experience
Knowledge formula	The tool that you just viewed connected lifestyle factors with clients' issues through a regression model. To what extent are you familiar with regression models?	Very familiar / Somewhat familiar / Not familiar
Knowledge formula 2	How many courses did you follow in which regression was discussed? (For example, mathematics or statistics courses in high school or at university.)	None / 1 / 2 / 3 / 4 or more
Knowledge formula 3	In the regression model below, can you indicate which predictor (A or B) has the most influence on the outcome of Y? Y = 5 + 3A + 2B + e (Note: if you don't know the answer, don't guess, but enter "I don't know")	I don't know / A / B / It depends on the unit and the variance of A and B
General remark	Do you have any questions or comments about the study, or about the tool?	Open question

Appendix B, The dashboard for the experiment

Figure 14. The Dashboard with Low Transparency

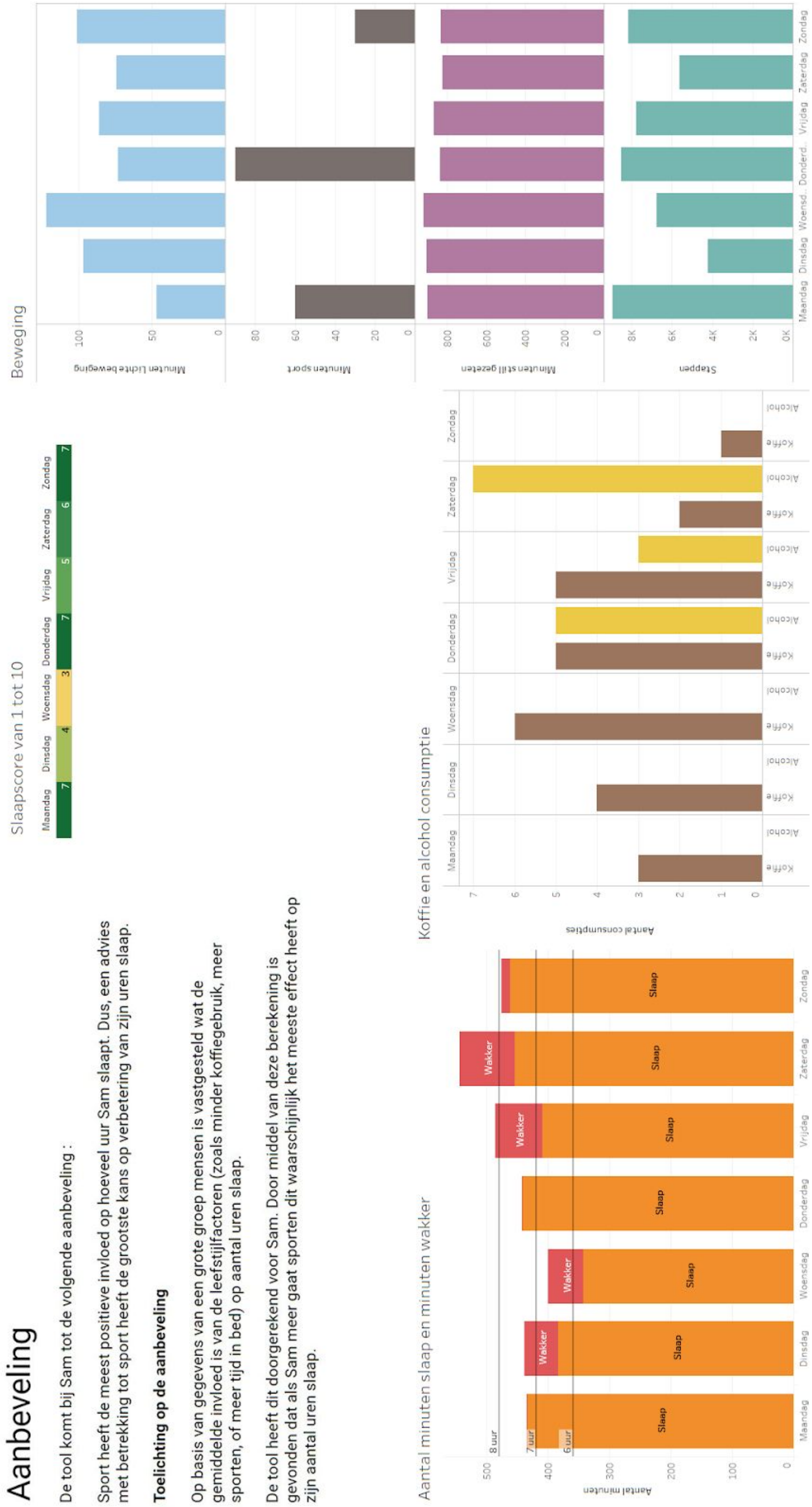

Figure 15 The dashboard with procedural transparency

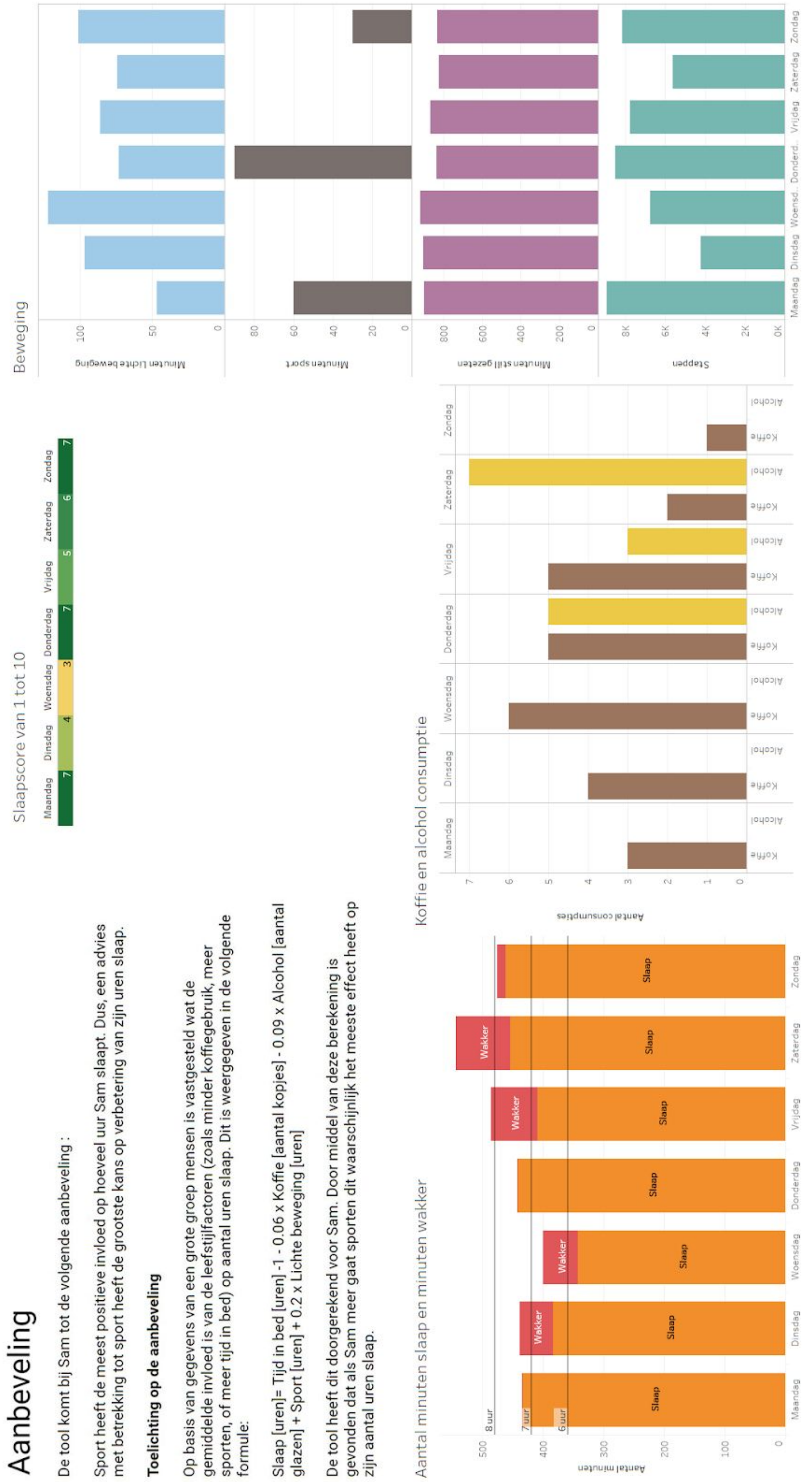

Figure 16 The dashboard with outcome transparency

Appendix C, Influential observations and outliers

First, the influential observations were inspected to discover if there were any participants with a big influence on the regression. Secondly, the outliers were inspected to see if there were participants which could be considered different from most other participants.

To check which observations had a considerable influence on the estimates of the parameters, the DFBETAs were inspected. The DFBETA measures the difference in each parameter estimate with and without the influential point. DFBETAs were considered large if they were bigger than the cutoff value $\frac{2}{\sqrt{n}}$ as proposed by Belsley, Hug, and Welsch (1980).

Table 10 shows the participants that had a large influence on at least five out of ten of the estimates. Out of the 14 participants listed in the table, seven saw the procedural transparency condition, and seven saw the outcome transparency, so they were equally spread over the conditions.

Table 10.
Participants with more than five DFBETAs which could be considered having a large Influence on the estimates.

Participant number	N of DFBETA > $\frac{2}{\sqrt{n}}$	Age	Work experience	Δ Trust
2	5	62	10	0.83
10	6	47	7.5	-0.50
12	8	35	10	-1.67
13	10	31	4	0.33
14	10	22	0.5	1.67
21	7	28	2	1.33
24	7	22	4	-1.50
32	10	23	2	0.49
33	10	19	2	1.17
34	10	19	0.5	-0.67
36	10	21	0	0.17
39	10	21	5	-1.00
83	6	28	7.5	0.67
106	7	21	5	1.50

The mean *Age* in this group was 28.5, with a standard deviation of 12.3, which is much higher than the overall mean and standard deviation (M = 23.1 years, SD = 6.5 years). The mean *Work experience* also was much higher compared to the overall mean (M = 3.64, SD = 3.64 vs M = 1.39, SD = 2.41). The mean for *Trust* is lower in this group compared to the whole group (M = -0.12, SD = 1.02 vs M = -0.01, SD = 0.51).

These findings indicate there are groups underrepresented in our sample. Mainly participants with more work experience. To test this Fisher's exact was used to see if the DFBETAs are equally distributed over *Work experience.* The p-value for the two-sided Fisher's Exact test is < 0.001; therefore, the assumption that the DFBETAs were equally distributed over *Work experience* could not be accepted. This shows the participants with high influence on the regression model are not equally distributed over work experience in our sample.

Overall, the influential observations were older, had more work experience, and had less understanding of regression models. These traits were the traits that were expected from more experienced health coaches. Because the data collection is not distributed equally over work experience, the more experienced participants were seen as influential observations with a strong influence on the estimates of the parameters.

Outliers

The data were reviewed to check for outliers. When checking for outliers, several measures for outliers were used. When a person has a *Trust*-value that is extremely different from the average *Trust*-value, we can check if the standardized score is higher than 3. When a person has a *Trust*-value that differs extremely from its predicted value, we can check if the standardized residual is higher than 3. When people score very high on X (so-called *hat element* or *leverage point*), we investigate whether their standardized predicted value is higher than 3 and we investigate if there are observations with leverage > 3 (k+1) /N, with k as the number of predictors and N as the sample size (Cardinali, 2013). When there are people that have a considerable influence on the estimates of other individuals, we check if *Cook's distance* is higher than 1 (Cardinali, 2013).

Standardized and studentized residuals

The standardized residuals were inspected. Nine cases were discovered with a bigger absolute standardized residual than three. The nine cases had an average *Age* of 23.6 years, with a standard deviation of 5.6 years and an average *Work experience* of 2.2 years with a standard deviation of 3.2. The *Age* is not that different compared to the whole group. However, *Work experience* is higher compared to the average of all observations. Because the studentized residuals measure the residual when the observation is not used to create the regression model, the studentized residuals are a better indicator of influential outliers. The same cutoff value was used for the studentized residuals as for the standardized residuals. Therefore, the studentized residuals were also inspected, and the same nine observations were considered to be outliers.

Leverage

Next, the leverage for each observation was investigated. Observations with high leverage have an extreme value in one of the predictor variables. The cutoff value for leverage is: 3(k+1)/n = 3(10+1)/111 ≈ 0.30. With this cutoff value, ten observations were identified with bigger leverage than 0.3.

Cook's distance

Finally, Cook's distance was inspected to find any influential observations on the predicted values for *Trust*. When an observation is bigger than 4/n, it can be considered an outlier. In this case, observations bigger than 4/n = 4/111 ≈ 0.04, have been inspected. Seven cases had been identified with a Cook's distance bigger than the cutoff value.

Inspecting the influential observations and outliers

When inspecting all the influential observations and outliers, seven participants were outliers on every outlier criteria (*Table 11)*. These observations also all had a large influence on all the estimates for the parameter (all 10 DFBETAs > 0.19). The mean *Age* is 22.3 (SD = 4.11), and the mean *Work experience* is 1.36 (SD = 1.41). These values were not considered different from the values for all the observations (*Age* (M = 23.1 years, SD = 6.5 years), *Work experience* (M = 1.39, SD = 2.41).

Table 11.
The Participants that were Outliers on all measures of Influence

Participant	Age	Work experience	Trust	DFBETAs	Condition
13	31	4	0.333	10	Procedural
14	22	0.5	0.166	10	Procedural
32	23	2	0.499	10	Outcome
33	19	2	1.166	10	Procedural
34	19	0.5	-0.666	10	Procedural
36	21	0	0.166	10	Outcome
39	21	0.5	-1	10	Outcome

Appendix C, The codes from the open question

Table 12.

The codes from the first open question.

Codes	Times this was mentioned	Example quote	Example quote 2	Example quote 3
stress/ mental health	24	“There is no information about Sam's feelings. For instance, bad sleep may be due to stress.”	“There might be other issues such as stress from his work/study/relation that could have a negative influence on his sleep. This information is crucial in giving solid advice, but is harder to gather in a number.”	“ Maybe there is something going on in Sams subconscious about for example trouble at work, or it could be genetic. This, however, can not be measured by any device.”
Sleep times	18	Perhaps at which time he goes to bed, because that might be important for his sleeping as well.	It would give more insight if Sam's bedtime	Information about if he needs to wake up early (eg. an alarm will obviously cut his sleep for short).
Food intake	7	I do miss the amount of sugar he eats during the day, but for the rest, the data is enough.	Some things that could also influence sleeping are stress or what Sam ate.	How much Sam has eaten on a day could be useful to know
Coffee and alcohol times	6	I feel like the information that is missing would be the time of day he drinks coffee or alcohol		
Type of exercise	6			
Sleep stages (REM, Deep	6	Totaal energieverbruik en meer informatie over kwaliteit van	Daarnaast ben ik benieuwd naar de kwaliteit van zijn slaap. Hoeveel	The different sleep stages are missing which show the sleep quality in particular

sleep, ect.)		slaap (diepe slaap, REM slaap etc)	minuten diepe slaap, hoeveel minuten wakker in de nacht, hoeveel minuten REM slaap of lichte slaap. Slaapt hij alleen of heeft hij kinderen die hem ook wel eens in de nacht wakker maken, of een vrouw die snurkt ;).	
Clients subjective sleep quality	4			
Environment	4	I would like to see his personal environment, factors from outside (city/village/eating behavior) that could influence his sleep.	What is the social context and environment in which Sam lives	
Screen time	4	Ja. Eventueel als aanvulling: beeldschermtijd	Bespreken wat factoren zie die slaap kunnen verbeteren op de dinsdag en woensdag (zoals beweging, screentime, licht, stress/ontspanning)	Also the time when he last saw white light, so whether he is using black screen mode etcetera.
Clients goals	3	Er zijn meer zaken die van invloed zijn op beweging en slaap. Naast de gegevens uit de tool zou ik bij Sam nog een aantal dingen uit willen vragen/willen weten; leeftijd, medische voorgeschiedenis, medicatie, fysiek	“het is niet bekend wat zijn doelen zijn, wat id zijn hulpvraag, wat is zijn probleem. Dus een gericht advies geven is niet mogelijk.”	“ Coachen op gedragsverandering heeft voor een heel groot deel te maken met motivatie. Ook zou ik graag Sam zijn doelen weten, ik zou zijn vitaliteit en stress willen meten (op basis van evidence based tools zoals bv de vita-16 van tno) en ik zou willen weten hoe fit

		en psychosociaal functioneren, woon/werksituatie, persoonlijke wensen/doelen.		Sam fysiek is.”
Heart rate	3			
Food times	3	information about the consumption of food and his schedule of the day, like when does he eat/drink or go to bed.	Maybe you could also include information about the consumption of food and his schedule of the day, like when does he eat/drink or go to bed.	
Fitness level	2			
Type of work	2			
Precise day by day movem ent	2	nee ik mis het beroep van de persoon. en hierbij de dagindeling en welke momenten de persoon sport en beweegt		
Water intake	1	voldoende vocht? (water, of niet caffeine/alcohol)		
Time spend outside	1	I would like to see the time spent outside, as this is a good indicator of mental health.		
Total energy use	1	Totaal energieverbruik en meer informatie over kwaliteit van slaap (diepe slaap, REM slaap etc)		
Amount of coffee	1			

and alcohol in ml				