MissFit: Towards Mediating Missing Data in Personal Informatics (PI) Systems

by

Nannan Wen

Bachelor, Beijing University of Civil Engineering and Architecture, 2015

Submitted to the Graduate Faculty of

the Department of Computer Science in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF COMPUTER SCIENCE

This dissertation was presented

by

Nannan Wen

It was defended on

March 27 2024

and approved by

Dr. Jacob Biehl, Associate Professor, Chair

Dr. Alexandros Labrinidis, Professor, Co-Chair

Dr. Erin Walker, Associate Professor

Dr. Dmitriy Babichenko (DINS), Clinical Associate Professor

MissFit: Towards Mediating Missing Data in Personal Informatics (PI) Systems

Nannan Wen, PhD

University of Pittsburgh,

In today's fitness-focused society, individuals with diverse technical backgrounds are increasingly engaging in the collection and analysis of their fitness data through movements like the Quantified Self (QS) and dedicated communities focused on Personal Informatics (PI) systems. However, these tools face a common challenge—they often struggle to handle missing activity data effectively, thereby limiting users' ability to gain profound insights. Researchers in the Personal Informatics (PI) and Human-Computer Interaction (HCI) communities suggest that PI tools should ubiquitously collect data. Despite ongoing efforts to optimize battery life, a fundamental issue persists: users must actively remove the device for charging, posing a risk of forgetfulness, especially during vacations.

This dissertation addresses the issue of missing data in Personal Informatics through four key contributions: 1) It explores existing algorithmic methods for handling missing data, constructs models for estimation, and evaluates their potential utility on a published Fitbit dataset. After this exploration, we discovered that the estimation models cannot make accurate estimates at an individual level, indicating a need for a new approach involving human-in-the-loop strategies. 2) We identify two distinct user groups and three primary usage behaviors from the semi-structured user study. These findings reveal that *maintainers* prefer to *know the present*, and *trainees* prefer *understanding the past*, and *predicting the future*. Missing data impacts trainees more than maintainers. Maintainers utilize data visualization and social features more on the PI tools, but trainees use data export and data analysis more. 3) We implement MissFit, a web app with three methods to handle missing data; these methods are algorithmic, event-based, and manual input; they are derived from natural approaches that fit with users' expectations discovered from the semi-structured user study. 4) After conducting an iterative user-centered design study, we identified four distinct user groups based on personalities and exercise routines: *SW*, *SP*, *UW*, and *UP*. People from different groups also present varying preferences for how to estimate missing data using different methods. All these contributions collectively lead to the proposal of six design implications. These implications

offer design principles on how future PI tools can utilize users' experiences to guide algorithms in estimating missing data and how to balance efficiency and user experience.

Table of Contents

List of Tables

List of Figures

Preface

The achievements in this dissertation were made possible thanks to the insights, encouragement, and support of numerous individuals. I faced a significant challenge due to a major change in my dissertation committee's leadership, and it nearly led me to consider abandoning my pursuit of a Ph.D. Fortunately, I was fortunate to have an understanding and supportive committee, department, family, and friends. Their collective efforts have enabled me to reach this point in my academic journey.

I am grateful to the department chair and my co-advisor, Labrinidis, for their generous support. I appreciate the opportunity to continue my Ph.D. journey under Labrinidis' guidance, especially after facing challenges in finding a suitable advising role within the department.

I also wish to express my gratitude to my advisor, Jacob Biehl, for his invaluable guidance in helping me refine my ideas, methodologies, and dissertation results, even before he officially took on the role of advisor. Professor Biehl has played a pivotal role in not only shaping my understanding of human-computer interaction but also offering valuable insights for my career planning. He has set a remarkable example in his interactions with others, demonstrating his ability to serve in various roles throughout my Ph.D. journey. Whether I needed direct suggestions when facing idea challenges, support when questioning predefined outlines, or constructive advice when navigating career choices, Professor Biehl has been a constant source of guidance. Additionally, his assistance in meeting deadlines to achieve all the milestones in completing my Ph.D. has been greatly appreciated.

I also thank Erin Walker and Dmitriy Babichenko for their willingness to serve on my Ph.D. committee and for providing valuable advice and feedback. Specifically, they both offered help and guidance when I needed the most. I'm forever grateful for their kindness and care for a student. Most of all, I appreciate their ability to provide a different perspective to my work.

Zak Risha deserves a lot of thanks for his help in implementing MissFit. He was instrumental in helping me tackle many technical challenges that were faced in this dissertation. In addition, Zak is one of the most warm-hearted people I have met outside my department.

I feel very lucky to have known a group of people in the department who are kind and sup-

portive. I have faced a lot of pressure when pursuing my Ph.D., and every time I feel down, they are there to cheer me up. They also help me sort through my thoughts when I'm doing research and give feedback when needed. They created a positive atmosphere for sharing ideas and helping each other to make things better.

Last but not least, I thank my family for trying their best to support me through my PhD. Even though they knew little about the challenges of studying abroad and getting a Ph.D., they were always there to support me, trying to give suggestions for me to keep going forward. I also extend my thanks to Yibo Wang, a special person in my life, as someone whose actions spoke louder than words; he used his unique ways to provide companionship during my difficult times.

1.0 Introduction

Wellness refers to the state of being in good physical, mental, and emotional health. It is a holistic concept encompassing various aspects of a person's life, including their relationships, work environment, and overall lifestyle. Wellness is not just the absence of illness or disease but is a state of optimal health and well-being. There are several dimensions of wellness, including physical, emotional, social, intellectual, occupational, and spiritual health. It is about taking a proactive approach to one's health and well-being and making choices that promote physical, mental, and emotional balance and harmony.

Physical health refers to the overall condition of the body and its ability to perform everyday activities and tasks without feeling fatigued or physical stress [13]. It includes a range of factors related to the body, such as body composition, cardiovascular and respiratory function, muscular strength and endurance, flexibility, and overall energy levels [5]. Good physical health involves maintaining a healthy body weight, following a balanced and nutritious diet, engaging in regular physical activity, getting adequate sleep, avoiding harmful habits such as smoking and excessive alcohol consumption, and managing stress effectively. It also includes regular medical check-ups and preventive health measures, such as vaccinations and cancer screenings, to detect and address any health issues early on.

In contrast, mental health refers to a person's overall psychological well-being and ability to manage emotions and cope with life's challenges healthily and positively. Again, it encompasses many factors, including thoughts, feelings, behaviors, and social interactions [51]. Good mental health involves having a positive outlook on life, feeling good about oneself, and having healthy relationships with others. It also involves the ability to manage stress, overcome challenges, and cope with difficult emotions in a healthy and constructive manner. Mental health issues can range from mild to severe and can include conditions such as anxiety, depression, bipolar disorder, schizophrenia, and eating disorders, among others.

Wellness encompasses a variety of factors, including proper nutrition, exercise, stress management, sleep hygiene, social connections, and emotional well-being. It emphasizes the importance of taking care of one's body and mind through healthy habits and lifestyle choices. Wellness is often viewed as a continuum or spectrum, with different levels of well-being that can be achieved through various practices and behaviors. It's a dynamic and ongoing process that requires continuous effort and attention to maintain optimal health and well-being. Physical fitness is essential for overall well-being, as it can impact mental and emotional health, social relationships, and overall quality of life. Regular physical activity, in particular, has been shown to have numerous health benefits, including reducing the risk of chronic diseases, stroke, and diabetes, improving mood and cognitive function and enhancing overall longevity [99]. It's crucial to prioritize and maintain good mental health as it can impact all aspects of life, including physical health, work and school performance, social relationships, and overall quality of life. Mental health can be improved and maintained through various strategies, including therapy, medication, exercise, self-care practices, and social support. Seeking help when needed is critical to maintaining good mental health and well-being.

That being said, in order to promote their physical, mental, and emotional well-being and enjoy a better quality of life, people engage in regular exercise and physical activity to help boost energy levels and to perform daily tasks more effectively and efficiently [24, 46]. In addition, many fitness activities, such as group exercise classes, team sports, and outdoor activities, provide people with opportunities for socialization and meeting new people. Overall, being fit can be a source of personal satisfaction and accomplishment if individuals are able to set and achieve fitness goals and improve their overall health and well-being.

In recent years, there has been a growing interest in health and fitness, with many people recognizing the importance of caring for their bodies and minds. Regular exercise, a balanced diet, and proper rest are all become crucial components of a healthy lifestyle. As a result, more and more people are trying to incorporate these practices into their daily routines. Furthermore, the report shows that the national health expenditure (NHE) grew 2.7% to $4.3 trillion in 2021, or $12,914 per person, which accounted for 18.3% of the Gross Domestic Product (GDP) [23]. A recent study also showed that Americans spend 11.7 hours per week on their health and fitness regime [14]. In addition, individuals who engage in more physical activities can reduce the risk of early death [74]. They are also known to be better at focus and more energic [83].

There are many ways that people can improve and maintain their fitness. One of the most common methods is regular exercise. It can include activities such as brisk walking, jogging, swimming, cycling, weightlifting, or group exercise classes. Strength training is another method that people utilize. It is an essential component of fitness, as it helps build and maintain muscle mass and improve overall strength and endurance. It can include lifting weights or using resistance bands. In addition, cardiovascular exercise is vital for improving heart health and endurance. This can include activities such as running, cycling, or swimming. Flexibility and mobility exercises, such as stretching or yoga, can help improve joint mobility and reduce the risk of injury. Eating a balanced and healthy diet is essential for maintaining good fitness. This can include consuming plenty of fruits and vegetables, whole grains, lean protein, and healthy fats. Rest and recovery are essential to wellness, allowing the body to repair and rebuild after physical activity. Getting enough sleep and allowing for rest days are important for maintaining overall fitness. Managing stress is important for general well-being and can also have a positive impact on wellness. This can include meditation, deep breathing, or taking a yoga class.

1.1 How Wearables Motivate People to be Healthy

In establishing healthy habits and making sustainable lifestyle changes over the long term. Individuals often leverage the usage of smart devices or applications to stay motivated and accountable. The history of wearable devices dates back to the 1960s when the first wearable device, a heart monitor, was invented. In the 1990s, companies began developing fitness trackers that could measure steps, distance, and calories burned, which became increasingly popular among fitness enthusiasts. The launch of the first Fitbit in 2009 marked the beginning of the wearable consumer revolution, which has since exploded with the introduction of smartwatches, smart glasses, and other types of wearables. Wearable devices now incorporate a range of sensors, such as accelerometers, gyroscopes, and barometers, to measure various aspects of the environment and the body. They use Bluetooth, Wi-Fi, or cellular networks to communicate with other devices, such as smartphones or cloud servers, where data can be stored, analyzed, and visualized. The increasing popularity of wearables has led to the development of dedicated platforms and ecosystems, such as Apple's HealthKit, Google Fit, and Samsung Health, which provide developers with tools and resources to create applications that integrate with wearable devices.

Many of the wearables and applications today use *challenge*, *progress*, *gamification*, and *social interaction* to motivate users to exercise more [16]. Table 1 illustrates the features implemented in current wearables and applications that aim to help individuals in setting realistic goals and track progress to stay motivated and accountable. *Challenge* includes activity, workout, distance, adventure, etc. Activity challenge encourages users to complete their daily activity goals, i.e., standing for at least one minute each hour. The workout challenge encourages users to complete a certain number of workouts in a specific time period, such as a week or a month. The distance challenge encourages users to complete a certain distance, such as walking or running a certain number of miles. The adventure challenge takes users on a virtual adventure, where they can explore new destinations and landmarks by completing workouts and hitting certain milestones. *Progress* tracking motivates users to stay motivated and focused by allowing them to see their progress in real-time. Most wearable devices enable users to track their progress over time. *Gamification* is a concept of using game elements in non-gaming context [32], and it's also widely used in wearables. For example, Garmin offers badges and rewards for achieving certain fitness milestones, i.e., users can earn virtual badges and rewards if they complete a certain number of steps. *Social interaction* allows people to share their goals with their family members or friends. These connections can provide a supportive and encouragement for them to achieve their goals [16].

In addition, Federica, a researcher at Statista, reported that the number of wearable devices had grown by 167% from 2014 to 2022 [60], indicating a rise in the use of wearable devices in the United States. Furthermore, Jacqueline, an editor based in Southern California, notes that nearly half (45%) of Americans already wear smartwatches regularly, and a significant majority (69%) would be willing to wear a fitness tracker. This suggests that the use of wearable devices is becoming more mainstream, and people are open to adopting these technologies to improve their fitness. Individuals also benefit greatly from the biometric data wearables collected. Many individuals (42%) who use fitness trackers have discussed this data with their doctors [31], which suggests that this information is reliable and can be useful in improving their health [42]. Studies also showed that wearing smartwatches has a positive correlation with exercising more frequently. According to Jacqueline, individuals who regularly wear smartwatches tend to exercise an average of 4.3 days per week, which is 43.3% more than non-users [31]. This could be because smartwatches provide users with personalized fitness tracking, reminders, and motivation to meet their goals. The availability of biometric data and progress tracking may also incentivize users to stay on track and continue exercising regularly.

Personal trainers or fitness coaches are also helpful in helping people maintain fitness, as they can provide guidance and support not only in creating personalized workout plans but also help with injury prevention and rehabilitation. Especially for individuals who are new to exercise, recovering from an injury, or looking to achieve specific fitness goals. However, the cost of a personal trainer or fitness coach can be expensive for some people. Many applications fill this gap by serving as virtual personal trainers. For example, Fitbit Coach offers personalized workout recommendations and video instructions. MyFitnessPal (see Table 1) provides personalized workout and meal plans. Coach to 5k (see Table 1) is a popular app designed to help beginners gradually build up their running endurance and reach the goal of running 3.1 miles. Golden Cheetah (shown in Table 1) is designed for cyclists and triathletes to help analyze and improve their performance, where one can share their data with their personal coach.

With so many powerful wearables and fitness applications, it is easier than ever for users to track and monitor their personal fitness data. This has led to a growing interest in the Quantified Self (QS) movement and Personal Informatics (PI) systems, which aim to use technology to help individuals better understand and improve their health and wellness. By tracking and analyzing data on their daily habits and behaviors, users can gain insights into their overall health and make more informed decisions about their lifestyles.

The term "Quantified Self" was coined in 2007 by Wired magazine editors Gary Wolf and Kevin Kelly. Q-Selfers often utilize basic tools such as spreadsheets and paper journals to record and analyze their personal data. They would manually input data on a daily or regular basis, such as their weight, exercise routines, sleep patterns, and food intake. As technology advanced, they began to adopt more advanced tools, such as wearable devices that automatically track biometric data (see Table 1), as well as specialized apps and software [47, 70, 6, 4] that can help with data analysis and visualization.

Personal informatics is a research field that explores how individuals can collect, analyze, and reflect on personal data to improve their lives. Extensive prior research has examined the use and utility of these technologies to improve behavior [20, 80, 63], understand differences in engagement and reflection on personal data [80], and build descriptive and predictive models of

Table 1: The table displays the features currently implemented.

Category	Name	Motivation and accountability features			
		Challenge	Progress	Gamification	Social interaction
Wearables	Apple Watch	●	●	●	●
	Favero Assioma	○	●	○	○
	Fitbit Watch	●	●	●	●
	Fossil Watch	●	●	○	●
	Garmin Watch	●	●	●	●
	Huawei Watch	●	●	●	●
	Kronos Watch	●	●	○	●
	Pedometer	○	○	○	○
	Power Meter	○	●	○	○
	Samsung Watch	●	●	●	●
	Suunto Watch	●	●	●	●
	TicWatch	●	●	○	●
	Veryfit Watch	●	●	○	○
	Wahoo Elemnt	●	●	●	●
	Whoop band	●	●	●	●
	Xiaomi Watch	●	●	●	●
Applications	Apple Health	○	●	○	●
	Couch to 5K	●	●	●	●
	FitOn	●	●	○	●
	Garmin Connect	●	●	●	●
	Golden Cheetah	—	—	—	●
	Headspace	●	●	●	●
	Lose It	●	●	○	●
	My Fitness Pal	●	●	○	●
	Nike Training Club	●	●	●	●
	Runtastic	●	●	●	●
	Seven	●	●	●	●
	Strava	●	●	●	●
	Sworkit	●	●	○	●
	Virgin Pulse	●	●	●	●
	Zombies, Run!	●	●	●	●

Note: The symbol "●" means that the feature is available in the wearables/applications, the symbol "○" indicates that the feature is not implemented, and the "—" symbol denotes that the wearables/applications do not possess a particular feature. The primary objective of these features incorporated in wearables/applications is to aid participants in sustaining their motivation and accountability.

Table 2: The table displays the data types collected.

Category	Name	Sensors	Workout types	Customizable
Wearables	Amazfit	Heart rate sensor,	Walking, Running, Biking, Cardio	yes
	Apple Watch	ECG sensor, Blood	workouts, Weight lifting, Yoga, Pilates,	yes
	Fitbit	Oxygen (SpO2)	Martial arts, Hiking, Dance classes,	yes
	Garmin	sensor,	Circuit training, High-Intensity Interval	yes
	Huawei Watch	Accelerometer,	Training (HIIT), Tennis, Golf,	yes
	Kronos Watch	Gyroscope,	Skiing/Snowboarding, Climbing,	no
	Samsung	Barometer, Ambient	Swimming, Strength Training, Indoor and	yes
	TicWatch	light sensor, GPS,	outdoor rowing, Elliptical workouts, Stair	yes
	Veryfit Watch	Barometer, Compass,	Stepping, Stand-up Paddleboarding	yes
	Xiaomi Watch	Thermometer	(SUP), Kayaking, Multisport activities	yes
			(such as triathlons)	
		Offers	**Recorded data types**	
Applications	Apple Health	Daily (weekly) summaries,	Steps taken, Distance traveled,	yes
	Coach to 5K	Insights, Progress tracking,	Active calories burned, Calorie	no
	FitOn	Goal tracking, Third-party	intake, Water intake, Sleep duration,	yes
	Garmin Connect	app integration, Community	Sleep quality, Heart rate, Blood	yes
	Golden Cheetah	support, Personalized	pressure, Respiratory rate, Menstrual	yes
	Google Fit	recommendations, Guided	cycle, Medical conditions, Walking	yes
	Headspace	meditation, Sleep aids,	speed, Walking distance, Active	no
	Lose It	Training plans, Personalized	time, Workouts, Body weight,	yes
	MyFitnessPal	Coaching, Data analysis,	Recipes, GPS location, Personal	yes
	Nike Training	Performance tracking, Race	bests, Rewards, Pace, Power,	no
	Club	analysis, Data sharing,	Cadence, Elevation, Training Stress	
	Runtastic	Personalized health coaching,	Score (TSS), Intensity Factor (IF),	yes
	Seven	Activity tracking, Health	Food intake, Workouts, Workout	yes
	Strava	content and resources, Social	history, Stroke count, Stress levels,	yes
	Sworkit	support, Challenges and		yes
	Virgin Pulse	competitions, Gamification,		no
	Zombies, Run!	Reminder and motivation,		yes
		Nutrient analysis		

Note: The table reveals a wide range of collected data by current wearables/applications, some of which overlap, while others differentiate themselves by focusing on specific activities. However, missing data remains a persistent problem across these devices. To address this issue, current apps provide a solution that allows users to input missing data manually, also known as customizable.

specific events and activities [62, 41]. Research has also sought to expand what can be quantified, exploring new use domains [18, 22, 39, 40, 86]. Even though self-tracking and personal informatics can be a powerful tool for individuals to gain insights into their behavior and make informed decisions about their health and well-being. One common pitfall in self-tracking is tracking too many things [20], particularly in the initial stages of exploration. Individuals may become excited and eager to track every aspect of their lives, but this can lead to using multiple tracking tools that capture different types of data (see Table 2), resulting in a large amount of data in different formats that can be difficult to analyze and interpret.

To avoid this pitfall, individuals start by identifying the specific areas they want to track and choose tracking tools that capture the relevant data in a consistent format [20]. Additionally, goal-driven tracking [84] is another effective strategy for individuals to stay motivated and track progress toward achieving specific objectives. For example, a person might use a fitness tracker to set a goal of taking 10,000 steps per day and use the data collected by the tracker to monitor their progress towards that goal. Others might be tracking food intake to manage weight, sleep to improve sleep quality, and weight lifting to build muscles.

But even with specific goals in mind, people often need to use multiple applications to track relevant variables. That being said, different applications need to be able to share data with each other. To address the issue of using multiple applications to track relevant variables, some companies are developing integrated platforms that allow users to track multiple aspects of their health and fitness in one place. For instance, some fitness tracking apps can sync with nutrition tracking apps to provide a more comprehensive view of a user's health and fitness. Despite the progress made in data integration, errors still occur, and this can lead to frustration among individuals. For example, a device may fail to record accurate data, or the data may not sync correctly with the user's preferred tracking app. This underscores the importance of data accuracy and the need for users to carefully evaluate the reliability of the devices and applications they use for tracking.

Battery life is another important consideration for wearable devices and self-tracking tools. With continuous use, these devices can quickly drain their battery, which can be inconvenient for users. In particular, if the device runs out of battery during a workout or activity, the user may lose all the remaining data they were tracking during that time. Additionally, forgetting to charge the device can lead to a missed opportunity to track activity and progress. While efforts have been made to improve battery life through software optimizations and the use of more efficient hardware components, the fundamental issue of having to take off the device and plug it into a charger remains. This inconvenience can be a significant obstacle for users who want to maintain a consistent tracking routine, especially for those who rely on their devices for daily tracking or athletic training. Therefore, unless the advancements in technology allow devices to charge seamlessly and on-demand, without the need for users to take them off or remember to charge them, battery life will continue to be a significant challenge for self-trackers.

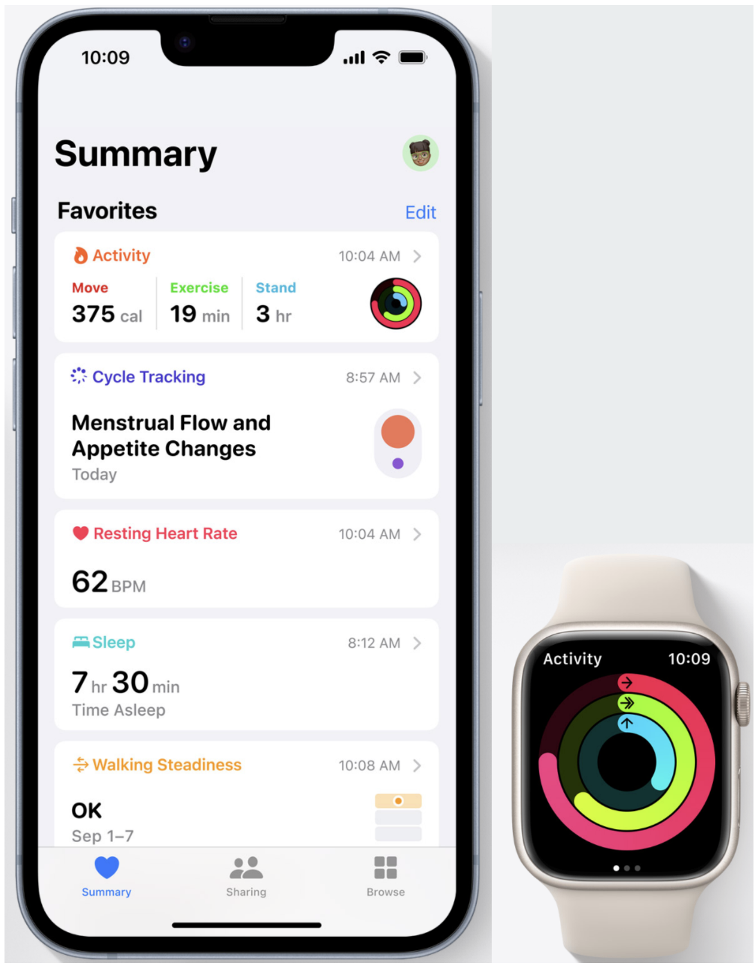
Take Alice, a busy working professional, as an example. Alice is dedicated to maintaining her fitness and health. She has been using a Fitbit device for the past few months to track her daily activity, including steps taken, calories burned, and sleep patterns. However, Alice often forgets to charge her Fitbit device, which causes her to miss out on tracking her activity and progress for days at a time.

For example, Alice wakes up early on Monday morning and goes for a run before work. She puts on her Fitbit device and starts tracking her run but soon realizes that her device has a low battery and needs to be charged. She doesn't have time to charge it before work, so she takes it off and puts it aside. Throughout the day, she forgets to charge her device and, as a result, misses out on tracking her steps, calories burned, and sleep that night.

This pattern of forgetting to charge her device continues throughout the week, causing Alice to miss out on valuable tracking data. She becomes frustrated with herself and her device, realizing that the battery life is a significant obstacle in maintaining a consistent tracking routine, and decides to abandon the device.

One potential solution to address the problem of missing data due to forgetfulness in charging their Fitbit device is to develop systems that can help individuals retrieve their missing data based on patterns and trends in the available data. This would provide a complete picture of people's daily activity and help them maintain a consistent tracking routine, even when they forget to charge their devices.

Figure 1: Current interactions between smartwatches and smartphones.



Note: In this scenario, the Apple Watch on the right stays on the individual's wrist and pairs with an application on the smartphone. Once the Apple Watch collects one's data, it will sync with the smartphone to upload it. If the individual forgets to sync it or the device malfunctions during the syncing attempt, missing data may occur.

1.2 Problems Missing Data Causes

The problem of missing data in personal data tracking has been recognized by several researchers, including those in Quantified Self-movement. Missing data can have several negative impacts, such as bias in analysis, leading to incorrect conclusions. It can also prevent users from gaining valuable insights into their behaviors [62], potentially leading them to stop tracking altogether [41]. Moreover, missing data can cause emotional distress to users, as they respond more positively to devices that capture accurate data overall [20]. Missing data can also cause users to mistake certain events as important or cause them to put in extra effort to ensure the device records their activities correctly [80]. Several studies have highlighted the negative impact of missing data, such as participants complaining about missed steps in their daily step count [26] or feeling terrible when there is a missing bar on the summary tab in menstrual tracking studies [40], and thinking that they were not getting proper credit for their activities[48].

Furthermore, missing data can reduce the accuracy and reliability of the data, which can undermine the trust users have in the tool [85]. Studies have shown that people abandon their wearable devices because of the quality of collected data [28, 38]. For instance, one in ten American adults owns some form of activity tracker, but half of them no longer use it [44]. Despite numerous efforts that personal informatics (PI) tools made to integrate into users' fitness routines seamlessly, missing data remains a persistent problem and a significant contributor to device abandonment [9, 43, 62, 20, 41]. Research by Attig et al. [9] found that 32% of participants cited "forgot to wear tracker" as a reason for abandoning their wearable device, while 27% cited "data synchronization problems." Another common reason for abandonment was a dead battery during a workout [43]. Other studies have reported that sparse or unhelpful data can deter users from adopting wearables [63].

Ultimately, missing data can affect the tool's effectiveness and influence users' usage, leading to the abandonment of the tool altogether [55]. These findings highlight the importance of addressing missing data and providing tools that can help users mitigate its impact on their tracking activities. As such, PI tools need to find ways to collect complete data sets and develop mechanisms to prevent data loss, such as data backup systems or reminders to wear devices. By improving the accuracy and reliability of data, these tools can increase user trust and motivate individuals to continue tracking their personal data for better health and wellness outcomes.

1.3 Research Questions and Summaries of Findings

- **RQ1**: What are the practical and fundamental limitations of fully automatic methods for mitigating missing data?

  Automated methods for estimating missing data involve utilizing statistical and machine learning algorithms to analyze available data and predict missing values. In our research, we investigated several existing methods and developed our regression-based model to estimate missing data in fitness tracking data.

  Despite the effectiveness of algorithmic-based approaches, we discovered fundamental limitations in these methods. For instance, unlike heart rate, running data is sporadic because individuals do not engage in running activities twenty-four-seven, making it challenging to distinguish between missing values and instances where there were genuinely no running activities. Another complicating factor is time, which is difficult to predict accurately. For example, a user might have missing data from 2:00 pm to 5:00 pm, but obtaining that information purely from an algorithm is challenging.

  Building on these insights, we believe it's essential to incorporate humans in the loop when estimating missing data. This process involves understanding the user requirements of the wider population using self-tracking tools and their experiences with missing data. By combining human expertise with automated algorithms in a human-in-the-loop strategy, we developed a system that effectively addresses the challenge of missing data in personal informatics.

- **RQ2**: What concerns about missing PI data are universal across users? What concerns are unique to specific populations?

  In order to gain insights into users' expectations of consistency and completeness in PI data, we conducted semi-structured interviews with participants who regularly engage in self-tracking activities. During these interviews, we asked participants about their past experiences with tracking, including what types of data they collect, what devices and tools they use, and their expectations for the completeness and accuracy of their data.

The analysis showed two distinct classes of users: ***trainees*** and ***maintainers***. Trainees have concrete, professional athletic training goals and use data to guide decisions to improve performance-related metrics. Maintainers have broad health improvement and maintenance purposes and use data to track milestones. We identified distinct user behaviors and perceptions when data was missing in personal informatics (PI) tools. Interestingly, both classes of users utilized the same tools in different ways. Trainees preferred data export and analysis functionalities, while maintainers favored visualization features. We captured the sentiment and motivation for these divergent behaviors in the interviews, with trainees primarily using their PI tools to understand their past and predict the future. In contrast, maintainers use them to know the present. Our resulting analysis points to key limitations of current tools, specifically the lack of representation for missing data. It outlines design observations for future tools to improve the user experience with PI data.

- **RQ3**: What techniques do users find useful and effective for eliciting missing data? What are the context or experience mechanisms used to support these processes?

  During our semi-structured interview study aimed at understanding the impact of missing data on end-users tracking activities, we specifically asked participants about their experiences with missing data, including how frequently they encountered it, how they perceived it in relation to their tracking goals, and how they coped with it.

  Our analysis revealed three usage behaviors of Personal Informatics (PI) tools: *understanding the past*, *knowing the present*, and *predicting the future*. We also found that current tools lacked support for integrating these behaviors, had no visual distinction between missing data and low activity, and lacked contextualized information about missing data. We found that visualizing missing gaps can provide an opportunity for users to estimate missing data manually or enable an algorithm to generate heuristically derived synthetic data. Additionally, we contributed to the understanding of the high trust users have towards collected data and how they often compare their data with that of other individuals, examine real-time maps, and estimate data based on their physical sensations to verify this trust further. Based on these findings, our study suggests that tools enabling synthetic data creation should allow users to observe synthetic data within the context of plausible and representative behaviors, group and label synthetic data as specific events, and allow users to manipulate and revise algorithmic estimates based on intuition or preference.

- **RQ4**: What designs and interactions are effective in supporting user-driven elicitation and estimation of missing data?

  The above study found that participants have varying levels of tolerance for missing data, with some accepting one or two days of missing data per week while others tolerate only a few hours per day. Participants expressed tolerance for missing data when it did not interfere with tracking goals, trends, and specific activities. The study suggests three design approaches for handling missing data: 1) *Manual-input method.* Allowing users to fill in missing data manually for small gaps. 2) *Algorithmic method.* Using algorithmic mediation to estimate gaps in data for large data gaps in daily, repeated activities. And 3) *even-based method.* Using algorithmic mediation and guiding user interfaces to help users refine estimates based on known conditions and events of the missing data for larger gaps.

  To better understand the effectiveness of each approach and the conditions under which users prefer each method, we conducted an iterative user-centered design study to gather feedback on the proposed system (MissFit) and its elicitation mechanisms. This study identified four distinct user groups based on personalities and exercise routines: ***SW*** (Structured, Wellness-oriented), ***SP*** (Structured, Performance-oriented), ***UW*** (Unstructured, Wellness-oriented), and ***UP*** (Unstructured, Performance-oriented). Individuals from these groups exhibit varying preferences for handling missing data. For instance, participants from groups UW and UP found remembering specific numerical values or activities challenging but found recalling their daily routines (i.e., work schedule) more manageable. Based on these findings, we proposed six design implications for how future Personal Informatics (PI) tools can leverage users' experiences and routines to guide algorithms in estimating missing data while balancing efficiency and user experience.

1.4 Existing Approaches

One approach to mediating battery dies during the workout is to improve the battery life of wearable devices. For example, the Apple Watch uses a low-power display that only refreshes when needed, a custom-designed system-in-package that integrates multiple components into a single chip, and an efficient power management system that automatically adjusts power usage based on usage patterns. Similarly, the Samsung Galaxy Watch uses a high-capacity battery, a low-power processor, and a power-saving mode that disables non-essential features when the battery level drops below a certain threshold. Other wearables, such as the Fitbit Versa and Garmin Forerunner, also use similar combinations of hardware and software optimizations to improve battery life.

Many wearables, including Apple Watch, Samsung Galaxy Watch, and Fitbit, allow users to adjust various settings as another way to improve battery life. Users can adjust the frequency of data synchronization, the brightness of the screen, and the use of certain features like GPS tracking. For example, the Apple Watch allows users to adjust the frequency of data fetching, turn off background app refresh, and enable power-saving mode to conserve battery life. The Samsung Galaxy Watch allows users to adjust the screen brightness, enable power-saving mode, and turn off GPS tracking when not needed. Fitbit also allows users to adjust settings such as the frequency of data synchronization, screen brightness, and GPS tracking to optimize battery life.

Some wearables may send reminders to help users manage their battery life. For example, the Fitbit app can send notifications when the battery is running low and remind users to charge their devices. The Apple Watch also has a low-power mode that can be activated when the battery is running low to help extend its lifespan. Some wearables may also allow users to customize their notifications or use vibrating alarms to remind them to charge their devices.

The methods mentioned above can help with improving battery life and making it easier for users to charge their devices. Still, they do not address the issue of people forgetting to wear their devices, which is the main reason of causing missing data. This can be a challenging problem to solve, as it is largely dependent on user behavior and habits.

As for the algorithmic perspective, the current approach to mediating missing data depends on the specific domain and context. In general, there are several common approaches, including:

- Imputation: Fill in the missing data with estimates based on the available data.
- Ignoring missing data: some statistical analyses can simply exclude any observation with missing data.
- Multiple imputation: generating multiple imputed data sets, where missing values are replaced

by plausible values that reflect the uncertainty of the imputation process.

- Machine learning-based methods: using machine learning techniques to predict missing values based on available data.

The approach used depends on the specific research question and the context in which the data is collected. It is important to choose an appropriate approach to minimize bias and maximize the validity of the results.

To the best of my knowledge, no PI tools have considered incorporating missing data elicitation in their systems yet.

1.5 Scope

Goal-driven fitness self-trackers can be divided into two main groups: ***maintainers*** and ***trainees***. ***Maintainers*** are individuals who aim to maintain their health and often prioritize activities such as exercise, a balanced diet, stress management, sleep, and hydration. To track their progress in these areas, they may use different tools such as fitness trackers, nutrition apps, stress management apps, sleep trackers, and hydration tracking apps. For example, a person may use a fitness tracker to monitor their workout times and set a goal to run three times per week. Similarly, they may use a nutrition app to track their daily food intake and ensure that they are consuming a balanced diet.

On the other hand, ***trainees*** are individuals looking to improve their athletic performance and may have different tracking requirements. They may need to track specific metrics such as distance, speed, and training intensity. As well as monitor their heart rate variability, recovery time, and nutrition intake. They may use specialized tracking tools such as GPS running watches, heart rate monitors, and nutrition-tracking apps to achieve their goals.

Research has shown that the vast majority of people who use self-tracking tools and wearables are more interested in maintaining their health rather than training for specific athletic goals. According to a study conducted by the Pew Research Center in 2018, 60% of U.S. adults who track their health and fitness say they do so to maintain their overall health, while 16% say they track to improve a specific health condition or medical issue. Only 9% of respondents said they track to train for a particular athletic event or competition [76]. Therefore, in this thesis, we try to build systems that interest the ***maintainers*** most, as it means that the target user group is likely to be more interested in maintaining their overall health and well-being than in training for specific athletic events or competitions.

1.6 Our Solution

To develop a system to help individuals retrieve their missing data, we plan to use a human-in-the-loop strategy combining automated algorithms and human expertise. This approach involves developing a system that utilizes multiple mechanisms to elicit missing data.

The first mechanism is an *algorithmic approach* that utilizes mock-ups of a state-of-the-art imputation method to infer missing data. This approach involves finding an algorithm that can analyze the available data and use statistical models to estimate the missing data points, mocking up the generation of the estimated data with ground truth. The algorithmic approach has the advantage of being able to work quickly and efficiently.

The second mechanism is an *event-based approach* that utilizes historical event patterns to infer missing data. This approach involves analyzing a user's past events and using them to predict what activity they might have been doing during the missing data points. For example, if a user has a regular daily routine, this information can be used to infer what they might have been doing during the missing data points. This approach has the advantage of being able to utilize context-specific information to infer missing data.

The third mechanism is a *manual input approach* that relies on human memory to retrieve missing data. This approach involves prompting users to manually input their activities during the missing data points, either through a user interface or through a conversation with a human expert. This approach has the advantage of being able to retrieve highly accurate information for fairly recent activities, but it is limited by the time and effort required from the user.

By combining these three mechanisms, we can develop a robust and adaptable system for a wide range of scenarios. The *algorithmic approach* can quickly fill in missing data points based on statistical models, the *event-based method* can utilize contextual information to make more accurate predictions, and the *manual input approach* can provide highly accurate data through human expertise. The human-in-the-loop strategy allows for the development of a system that can fit a wider range of user populations.

This dissertation presents a new web app, MissFit, to help users elicit their missing exercise activities. A key technical innovation of MissFit is enabling users to get their missing data back under three scenarios: *algorithmic*, *manual input*, and *event-based* approach. Additionally, the prototype provides easy and intuitive interfaces that help guide users in getting their missing data back.

In refining MissFit, we followed an iterative user-centered design process, including surveys and interviews with current PI tool users. Through this process, we understood the common usage patterns users utilize within the context of PI tools. What motivations do they have, what information have they collected, and how do these tools help them achieve their goals? This understanding was pivotal in grounding the design of MissFit within the context of how users currently maintain fitness. Further, our work produced design implications that refine existing PI tools mediating missing data.

1.7 Contributions

This dissertation makes the following contribution:

- *Identification of the limitations and unique challenges related to missing data in the context of self-tracking activities.* Much of the existing work on missing data has largely focused on analyzing the overall dataset to obtain complete records. Although these methods can provide an instantaneous and effective approach to help end-users retrieve their missing data, they have fundamental limitations. For example, estimated missing data are often biased towards the dominant population and may not accurately reflect the personal characteristics and behavior patterns of end-users. Additionally, human behaviors can be unpredictable, making it challenging for machine learning algorithms to predict actions based solely on historical data. Our analysis emphasizes the need for more human-driven mechanisms, assisted by estimation technologies, to improve the overall experience and trust towards such a system (See Chapter 3 for more detailed discussion).

- *Identification of the human behaviors related to missing data in the context of personal informatics.* Existing research works have laid solid foundations for understanding user behaviors in personal informatics. Our work adds new insights into this field by identifying two distinct user groups: ***trainees*** and ***maintainers***, and three primary usage behaviors: *understanding the past*, *knowing the present*, and *predicting the future*. Our findings reveal that trainees often use PI tools to understand the past and predict the future, but maintainers often focus on knowing the present (see Section 4.2 for more detailed discussion). We also identified four distinct user groups based on personalities and exercise routines: ***SW*** (Structured, Wellness-oriented), ***SP*** (Structured, Performance-oriented), ***UW*** (Unstructured, Wellness-oriented), and ***UP*** (Unstructured, Performance-oriented). People from different groups showed different preferences when estimating missing data (see Section 6.2 for more details).
- *Identification of three missing data elicitation mechanisms and implementation of MissFit for estimating missing data.* We derived three elicitation mechanisms to facilitate missing data from natural behaviors: 1) Algorithmic method: Using estimation models to estimate missing data based on prior data collected; well-suited for larger data gaps in daily, repetitive activities, like estimating the number of steps taken during a routine evening walk. 2) Manual input method: Allows users to input missing data manually, which is ideal for small data gaps that correspond to specific activities, such as outdoor biking activity; one scenario is people refer to their friends' data when the data is missing. 3) Event-based method: Combines algorithmic estimation with user inputs to refine estimates based on known conditions of the missing data. It is useful for people who have regular exercise routines, and the missing data can be indicated by other activities or sets of activities. For example, estimating the running distance by considering the calories, neighborhoods, and GPS routes (see Section 4.3). Based on the findings, we implemented a web app called MissFit, which provides three estimation methods described above to help estimate missing data. Chapter 5 provides a detailed discussion of the implementation of MissFit.
- *New design insights for supporting different groups in mediating missing data in personal informatics systems.* Building upon all the lessons learned from this research activity, we proposed six design principles on how future PI tools can utilize users' experiences to guide the process of estimating missing data: 1) Utilize users' daily habits to assist in recalling missing

data. 2) Incorporate users' daily routines to estimate their data accurately. 3) Prioritize cumulative measurements while offering details of individual discrete activities. 4) Utilize users' real-time feedback to identify and estimate missing activities. 5) Aid users in understanding estimation methods by providing a complete and detailed summary of the "source" activities referenced. 6) Strike a balance between comprehensiveness and ease of use when estimating missing data. These design implications provide guidelines focusing on different user-centered design aspects (see Chapter 7 for detailed discussions).

2.0 Related Work

Personal Informatics (PI) is also known as "living by numbers," "quantified self," "self-surveillance," "self-tracking," and "personal analytics." It is defined as "tools that help people collect personally relevant information for the purpose of reflection and gaining self-knowledge" [62]. The Quantified Self (QS) movement played a vital role in popularizing the idea of self-tracking and personal data collection. Proposed by Gary Wolf and Kevin Kelly in 2007 through Wired magazine [104], the QS movement has gained popularity since then, encouraging individuals to use technology to collect data about themselves for self-reflection and improvement.

Even before the formalization of the QS movement, there were early examples of personal informatics tools. Platforms like Nike+ [102], launched in 2006, allowed runners to track their activities and performance, laying the foundation for developing more comprehensive PI systems later. Fitbit Tracker [101], released in 2009, marked a significant milestone in the history of Personal Informatics by bringing a wearable device to the mainstream. It not only tracked steps but also various health metrics, providing users with a holistic view of their physical activity and well-being. The Apple Watch [100], introduced in 2015, is considered one of the most popular and successful smartwatches in the market. It provides features Fitbit has and beyond, like receiving notifications for calls, messages, emails, and other app alerts on the watch, believing that a watch solely dedicated to fitness might not appeal to everyone.

As the use of personal tracking devices increased, researchers and academics delved into the implications of Personal Informatics. Studies investigated user experiences, motivations, and challenges associated with self-tracking, leading to a deeper understanding of the field. For example, Li et.al. [62] and Epstein et.al [41] modeled users' tracking behavior into different stages (see Section 2.1 for further discussions). Rapp et.al. [80, 81] explored challenges and designed solutions for users without tracking experience. The widespread adoption of smartphones and app ecosystems facilitated the use of health and fitness apps. Platforms like MyFitnessPal, Runkeeper, and Strava empower users to track various aspects of their health and wellness (see Section 1.1 for more details).

As PI tools became more prevalent, there was a growing emphasis on design and user experience. Researchers and developers started recognizing the importance of creating user-friendly interfaces and considering the psychological aspects of self-tracking. Personal Informatics became a distinct research area within the broader HCI field. Scholars delved into topics such as data visualization, behavior change, privacy concerns, and the impact of self-tracking on individuals' well-being. For example, Choe et.al. [19] how people reflect and gain insights through visual data exploration. Walsh et.al. [98] and Consolvo et.al. [27] studied how systems can intervene to increase users' physical activities (see Section 2.4 for more discussions). Missing data was identified as an important problem that hinders users from incorporating PI tools into their lives and leads to abandoning the tools (see Section 2.2 for more discussions). In this chapter, we present the research findings that are related to our work and how our work contributes to the field.

2.1 Human Behaviors in Personal Informatics

Human behaviors in Personal Informatics (PI) refer to how individuals engage with and respond to self-tracking tools, the motivations driving their tracking activities, and the impact of personal data on their behavior and decision-making. Several studies have been conducted to identify human behaviors in the context of PI.

The stage-based model in Personal Informatics, proposed by Li et.al. [62], is a conceptual framework that describes individuals' progressive phases when engaging with self-tracking and personal data. This model provides insights into the user experience and behavioral patterns associated with the adoption and continued use of PI tools. In the stage-based models, Li et.al. [62] divided the process of personal informatics into five iterative stages:

- **Preparation:** This stage involves the initial consideration and preparation for engaging in self-tracking. Users may express curiosity, interest, or a desire for self-improvement. They begin to explore available tracking tools and consider incorporating tracking into their lives.
- **Collection:** In this stage, individuals actively start collecting data about various aspects of their lives. Users begin tracking specific behaviors, activities, or metrics using devices, apps, or other tools. The focus is on gathering data that aligns with their goals or interests.

- **Integration:** In this stage, the collected information is prepared, combined, and transformed for users to reflect on. Users begin organizing the data collected from various tools or apps, with a focus on merging their data into one place.
- **Reflection:** During this stage, individuals reflect on the collected data to derive insights and meaning. Users analyze their tracked data to understand patterns, trends, and correlations better. Reflection may lead to behavioral adjustments or re-evaluation of goals.
- **Action:** Action involves using the insights gained from reflection to inform decision-making or goal re-evaluation. Users take proactive steps based on the information derived from their data. This could involve setting new goals, adjusting routines, or making lifestyle changes.

Epstein et al. [41] proposed the lived-informatics model, which considers the dynamic and evolving nature of individuals' lives. This model highlights the need to design Personal Informatics (PI) systems that can accommodate changes in users' routines and priorities. In addition to the stage-based model, the lived-informatics model further identifies a lapsing stage indicative of gaps in the usage of informatics. This stage is divided into four categories: forgetting, upkeep, skipping, and suspending, which are caused by various reasons and may or may not be resumed at a later time.

Both of these models laid a solid foundation for understanding user behaviors in personal informatics. Other researchers have also made efforts to comprehend the relationships between user experience and behavioral patterns with technologies. For instance, Swan et al. [94] delved into the potential of wearable devices and sensor data in shaping personal behavior and decision-making. Consolvo et al. [27] explored how users integrated self-tracking into their daily lives and the challenges associated with long-term engagement. Li et al. [63] emphasized the role of personal data in fostering a deeper understanding of oneself.

Our work contributes to understanding human behaviors in personal informatics by identifying two distinct user groups: ***trainees*** and ***maintainers*** (see Section4.2). It also identified three primary usage behaviors: *understanding the past*, *knowing the present*, and *predicting the future* (see Section 4.3.1). Different groups showed a distinct emphasis on usage behaviors. For example, the *maintainers* care more about knowing their current status, like how many steps they are at and how far they are from their daily or weekly goals (see Section 4.2.2 for more details), which has brought new insight into the reflection stage. Our work also found that different groups utilize the PI tools differently. For example, the *trainees* use the data export and data analysis aspect of the PI tools more often (see Section 4.2.1 for more details), providing new understandings about individuals' behaviors in the reflection stage.

In addition, we also identified four groups based on personalities and exercise routines: ***SW*** (Structured, Wellness-oriented), ***SP*** (Structured, Performance-oriented), ***UW*** (Unstructured, Wellness-oriented), and ***UP*** (Unstructured, Performance-oriented). People from different groups showed different preferences for how to estimate missing data. Group UW and UP were better at recalling their daily routines (i.e., class schedules) to inform the missing data, and group SW and SP were better at remembering trends or even specific numerical values to estimate the missing data (see Section 6.2 for more details). Our work collectively advances the understanding of tools and functionalities individuals in different groups prefer on their goal-tracking journey.

2.2 Missing Data in Personal Informatics

Missing data in Personal Informatics (PI) presented a significant challenge in accurately capturing and reflecting individuals' activities and behaviors. Users encountered practical obstacles that affected their ability to *collect* and *reflect* upon data using PI tools. The current literature has identified three main barriers to the data collection stage:

- Precision: the device itself is not precise enough [61]. The result shows a significant difference when people compare two different devices for the same type of sensory data, and sometimes the device malfunctions. Sometimes, data rely on the estimation of the user [62] (e.g., how many calories are in a specific meal).
- Remembering: when it is human-collected [41, 62], recall can affect the data collection. (e.g., people sometimes forget to record or do not remember what they did).
- Ubiquity: the device is not around when big moments happen [41, 62], which could be because the device needs maintenance (e.g., charging) or forgetting to wear. [or doesn't fit the environment (can not bring to the swimming pool, examples: functionality, sometimes the capable of doing something, add some examples)]

In the reflection stage, challenges arose from the sparsity of data and the difficulty in interpreting data, leading to difficulties in comprehending summaries, visualizations, and recommendations.

The mentioned obstacles have significantly impacted adherence, interest, and trust in PI [41, 26]. For instance, Choe et al. [20], who investigated the practices of the Quantified-Self movement, reported that participants valued tools (devices) capable of capturing comprehensive, granular information about their activities and expressed frustration when the collected data was inaccurate. Rapp et al. [80] found that missing data could lead users to mistake which activities were captured. In a study of long-term wearable device use, Thomas et al. [48] described that participants put extra effort into ensuring they had their tools (devices) with them before workouts, worrying they wouldn't receive proper credit for their activities without the devices. The identification of incomplete data can also impact users' affect; Epstein et al. noted that participants felt guilty when a menstrual tracking application's interface highlighted missing data [40].

The literature consistently highlights the pivotal role of missing data in shaping the value of Personal Informatics (PI) [20, 80, 48, 26, 40], with some researchers proposing solutions to address this challenge. For instance, Lazar et al. [61] recommended automating the data collection process by automatically detecting activities and minimizing maintenance by enhancing battery life.

Our work contributes to a deeper understanding of the conflicts arising between individuals' expectations and PI tools when encountered with missing data, outlining how these conflicts are both unique and shared between ***trainees*** and ***maintainers*** (see Section 4.2.2 for more details). We discovered that maintainers and trainees exhibit different tolerance levels towards missing data. While both groups reported frustration when encountering missing data, the magnitude of frustration differed. Trainees expressed that missing data had a more significant impact on their goals than on maintainers. Despite this, both groups demonstrated high trust in their tracking data, acknowledging the existence of missing and inaccurate data. We also identified some natural forms individuals employ to address missing data, such as writing it down, referring to friends' data, estimating based on how their body felt, or checking their prior routes.

2.3 Systems That Support Reflection in Personal Informatics

Systems that support reflection in Personal Informatics (PI) play a crucial role in helping individuals derive meaningful insights from their collected data. Reflection involves the interpretation, understanding, and learning from personal data to inform decision-making and behavior change [11].

Researchers have been focused on different aspects of reflection when building their systems. Some researchers tried to build systems that support reflection-in-action, which provides feedback during the activity. For example, commercial products like Fitbit and iWatch give real-time feedback when people do exercises or engage in activities so that they can monitor their vital signs. Breakaway [54], UbiFit Garden [27], and BeActive [21] were designed to provide interventions that encourage people to have a healthier lifestyle. The Ally+ app [71] acts as a chat-based digital coach to deliver in-the-moment interventions to motivate participants to achieve their step goals. The mHealth4U [89] aims to discover more targeted health and well-being self-management by increasing consumers' continuous engagement.

Some researchers tried to build systems that support reflection-on-action and provide feedback after the activity ends. For example, Fish'n'Steps [64] was designed to encourage physical activities by linking a player's daily step count to the growth of a virtual character, and the UbiFit [25] tried to achieve the same goal by using positive reinforcement based on past behaviors. The Visualized Self [19] devoted their effort to supporting deeper-level self-reflection through multiple data streams and visual data exploration using participants' historical data.

Some researchers tried to build systems to facilitate reflection through visualizations. A specialized field of "personal visualizations" aims to present personally relevant information that promotes actionable insights and subsequent changes in behavior. For example, Habito [49] focused on designing visualizations to increase awareness and encourage behavior change in self-trackers. Many of the systems provided visualizations in the form of dashboards and supported simple interactions to explore the data, employing timeline metaphors to present events chronologically. For instance, Lifestreams [52] was designed to extract specific behavioral indicators and inferences from linear, interactive visualizations. Moushhumi et al. [91] visualized time-series data to design just-in-time adaptive stress management interventions.

These systems facilitate reviewing and interpreting personal information, fostering self-awareness and informed decision-making. By providing tools for data visualization, summarization, and analysis, PI systems empower users to reflect on their behaviors, patterns, and progress. Additionally, these systems serve as crucial platforms for researchers to study the impact of different techniques on reflection, contributing to the advancement of how technologies can facilitate reflection.

Our work contributes to this field by implementing a web app called MissFit, which aims to facilitate reflection by assisting participants in recovering their missing data or activities (see Chapter 5 for implementation details). MissFit provides three approaches derived from natural behaviors to mediate missing data or activities:

- Algorithmic: This approach is designed to require minimal effort from users through automated algorithms.
- Event-based: This approach requires users to input information related to their missing event, helping the system estimate the missing data.
- Manual input: This approach depends on users' memory of the missing data.

Collectively, these three approaches address different user requirements and can accommodate changing behaviors (see Section 6.2 for more details). MissFit facilitates reflection by providing richer insight through the estimation of missing data or activities. For instance, by differentiating missing data from no data, MissFit helps users have a more accurate idea of what their day looks like. The approaches in MissFit also assist participants in recalling what happened to their missing data, promoting reflection.

2.4 Design Implications for Personal Informatics Systems

The design implications provide guidance for developers and designers to create PI tools that better meet the needs and expectations of users while addressing challenges inherent in personal data tracking and reflection.

Researchers have derived design implications for self-tracking tools based on individuals' experiences. For instance, Choe et al. [20] proposed design implications to address common pitfalls for Quantified-Selfers in their tracking experience, such as tracking too many things, not tracking triggers and context, and lacking scientific rigor. These implications include providing early feedback to help identify what to track, supporting self-experimentation through design, maximizing the benefits of manual tracking, and promoting self-reflection. Epstein et al. [37] put forth design opportunities to assist self-trackers in identifying meaningful and actionable findings, incorporating contextually aware feedback and supporting transitions between phases. Bentley et al. [12] recognized the need for reminders when new observations became available. All these works collectively contribute to a deeper understanding of the self-tracker's needs.

Some researchers have proposed design implications for self-tracking tools with a focus on visualizations. For instance, Cuttone et al. [30] put forward four heuristics:

1. Make data interpretable at a glance.

2. Enable exploration of patterns in time series data.

3. Enable the discovery of trends in multiple data streams.

4. Turn key metrics into affordances for action, using visualization to facilitate reflection.

These heuristics provided guidelines for the design of systems that can facilitate reflection in self-tracking personal informatics. Guided by these heuristics, Cuttone et al. [29] deployed their mobile application to provide personal data on mobility and social interactions through interactive visualization interfaces for educational use. Eun et al. [17] further characterized types of visualization insights as detail, self-reflection, trend, comparison, correlation, data summary, distribution, and outlier, bringing new insights into how visualization can help individuals reflect on their tracking data.

Some researchers have proposed design implications aimed at promoting behavior change. For instance, Consolvo et al. [26] put forward theory-driven design strategies for persuasive technologies to assist individuals in changing their everyday behaviors. These implications include:

1. Abstract and Reflective: Utilize data abstraction to display information, encouraging users to reflect on their behaviors.

2. Unobtrusive: Present and collect data in an unobtrusive manner.

3. Public: Display and collect data that users are comfortable sharing, even if others may be aware of it.

4. Aesthetic: Ensure the design is comfortable and attractive, aligning with the user's personal style.
5. Positive: Incorporate rewards to encourage desired behavior [48, 68].
6. Controllable: Allow users to edit recorded activities. An example of a mobile display, UbiFit Garden [27], was developed following these design implications to encourage physical activity.

These design implications provided guidance in shaping the implementation of personal informatics systems that effectively support users in collecting, reflecting on, and making use of personal data from different aspects. For example, Choe et.al. [20] focused on user-centered design, Cuttone et.al [30] focused on data visualization, and Consolvo et.al [26] focused on reflection support.

Our work contributes to this field by proposing six design implications for future Personal Informatics tools aimed at designing systems to mediate missing data (see Chapter 7 for details). These implications focus on guiding systems in handling missing data effectively. The implications are as follows:

1. Utilize users' daily habits to assist in recalling missing data.
2. Incorporate users' daily routines to estimate their data accurately.
3. Prioritize cumulative measurements while offering details of individual discrete activities.
4. Utilize users' real-time feedback to identify and estimate missing activities.
5. Aid users in understanding estimation methods by providing a complete and detailed summary of the "source" activities referenced.
6. Strike a balance between comprehensiveness and ease of use when estimating missing data.

These implications address various facets of user-centered design. For instance, the first, second, and fourth implications leverage user experiences to facilitate the estimation of missing data. The third and fifth implications focus on addressing users' emotional needs to establish trust between the system and users. The sixth implication emphasizes accommodating users' preferences. Collectively, these six design implications advance the field by incorporating effective methods for mediating missing data, offering a more holistic approach to designing Personal Informatics tools that support a comprehensive understanding of users' personal data.

3.0 Analysis of the Algorithmic Approach for Estimating Missing PI Data

To answer the first research question (Section 1.3), we investigated several existing methods for dealing with missing data, developed our own regression-based model to estimate missing data, and applied that method to a Fitbit dataset. This dataset can offer insights into the behavior and habits of Fitbit users, as well as provide a useful testbed for evaluating algorithms and models designed to analyze and process personal tracking data. Additionally, the dataset was generated through a distributed survey, which makes it a good source for a diverse range of users, thus, a valuable resource for studying the use of personal tracking devices across different populations.

The seriousness of missing values in a dataset depends greatly on how much data is missing, the pattern of the missing data, and the mechanisms of the missingness [56]. Various methods can handle missing values, including deletion of the missing instance and substitution of the missing value with approximated values [36, 97, 67], a technique known as imputation. Several imputation methods have been proposed to handle missing values [36, 97, 67, 50], including mean substitution, median substitution, last value carried forward, regression, and neural networks. In some cases, a combination of multiple imputation methods, also known as multiple imputation, is applied to overcome the weaknesses of traditional imputation techniques. However, it is important to note that only a suitable solution will result in a robust design and sound analysis of the dataset, especially since PI tools collect high-dimensional and complex behavioral data. Moreover, such data analysis depends not only on attribute selection but also on personalized characteristics. Therefore, the method used to handle missing data is crucial, as improper handling can lead to incorrect insights drawn from it and give users wrongful information about themselves.

Chapter 3 highlights that missing data in personal informatics (PI) tools can take various forms, such as inaccurate or completely missing data points. Given that detecting inaccuracies in data points can be a more challenging task for automatic algorithms without human intervention, our initial efforts to address missing data in PI tools focused on understanding when and how data got missing. In this section, our goal was to provide insights into mediating missing data using an algorithmic approach to detect and estimate missing data automatically. More specifically, what are the practical and fundamental limitations of fully automatic methods for estimating missing data? To achieve this, we combined techniques from existing literature on missing data with the unique characteristics of PI devices. We developed a lightweight model for estimating missing data that can be applied to wearable devices, and analyzed the effectiveness and limitations of such an approach.

3.1 Methods

There are three common approaches to handling missing values: 1) deleting samples with missing data, 2) ignoring missing values in the analysis, and 3) imputing or filling in the missing values. Sample deletion involves removing any samples with missing data before training machine learning models. However, this approach can result in insufficient training data and a loss of insight into PI systems due to a reduction in the dataset size. Ignoring missing data can lead to biased and inaccurate conclusions [79]. Data imputation is a technique that replaces missing values with plausible values in time-series data. This method is preferred in various disciplines, including biomedical research [95], network analysis [57], education [75], and longitudinal studies [53].

Various techniques have been proposed by researchers to impute missing data using statistical, machine learning, and neural network methods. Statistical imputation involves using statistical estimation to replace missing data with known data, such as using mean or median values. Traditional machine learning-based imputation methods include Maximum Likelihood Expectation Maximization (EM) imputation [92], K-Nearest Neighbor (KNN) based imputation [90], and Matrix Factorization (MF) [96]. However, most of these methods do not adequately consider the temporal information in time series data. Deep learning methods such as Recurrent Neural Networks (RNNs) and deep neural networks models like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) have been proposed as alternatives to fill missing values in sequential data. Additionally, some researchers have extended hybrid architectures of Generative Adversarial Networks (GANs) to improve imputation performance. While these methods have shown promise in predicting missing values or patterns [105] in wearable device data, we argue that they may not be sufficient for our research goals for the following reasons:

- The imputation methods mentioned may not adequately capture the complex and dynamic nature of physical activity data captured by wearable devices, particularly for longitudinal studies.

- Some methods may not effectively handle missing data that are not missing completely at random (MCAR), such as intermittent missingness or non-ignorable missingness, which can be common in behavioral datasets.
- Some deep learning-based approaches may require large amounts of data and computing resources to train, which may not be feasible for small or resource-constrained studies.
- Insufficient amount of training/testing data is a universal fact. Deep neural networks necessitate large quantities of input data during the training phase, which is not realistic in the case of personal PI data.
- The behavioral data has a lower density compared to biometric data due to larger gaps and sporadic behavior in the user's tracking habit.

This section focuses on investigating regression-based machine learning models for imputing wearable-sensory time series data, as proposed by [65]. Our approach considers activity patterns, contextual information, and personal characteristics as covariates. To our knowledge, there are few existing studies that combine personal characteristics and time series using regression models for limited wearable data. Our evaluation is conducted on a multi-model sensor time series dataset from Fitbit.

3.2 Description of Dataset

The Fitbit dataset used in this study was obtained from Kaggle [73], an online community platform that allows users to find and publish datasets. The dataset includes minute-level output for physical activity, heart rate, and sleep monitoring from 33 Fitbit users who consented to the submission of their personal tracker data over a period of 31 days. As shown in Table 3, the dataset is segmented into several tables with different aspects of the data from the devices, providing lots of details about the user behavior [59, 7]. However, the heart rate and weight document contain only a small sample of seven and eight unique users, respectively, so they were not considered in the analysis.

Table 3: The table illustrates the Fitbit dataset.

Document Name	Type	Description
dailyActivity_merged	CSV	Daily activity over 31 days, including steps, distance, intensities, and calories.
dailyCalories_merged	CSV	Daily calories over 31 days.
dailyIntensities_merged	CSV	Daily intensity over 31 days. Measured in minutes and distance, dividing groups into four categories: sedentary, lightly active, fairly active, and very active.
dailySteps_merged	CSV	Daily steps over 31 days.
heartrate_seconds_merged	CSV	Exact day and time heart rate logs for seven users.
hourlyCalories_merged	CSV	Hourly calories burned over 31 days.
hourlyIntensities_merged	CSV	Hourly total and average intensity over 31 days.
hourlySteps_merged	CSV	Hourly steps over 31 days.
minuteCaloriesNarrow_merged	CSV	Calories burned every minute over 31 days (Every minute in a single row).
minuteCaloriesWide_merged	CSV	Calories burned every minute over 31 days (Every minute in a single column).
minuteIntensitiesNarrow_merged	CSV	Intensity counted by minute over 31 days (Every minute in a single row).
minuteIntensitiesWide_merged	CSV	Intensity counted by minute over 31 days (Every minute in a single column).
minuteMETsNarrow_merged	CSV	Ratio of the energy you are using in physical activity compared to the energy you would use at rest. Counted in minutes.
minuteSleep_merged	CSV	Log sleep by minute over 31 days. The value column is not specified.
minuteStepsNarrow_merged	CSV	Steps tracked every minute over 31 days (Every minute in a single row).
minuteStepsWide_merged	CSV	Steps tracked every minute over 31 days (Every minute in a single column).
sleepDay_merged	CSV	Daily sleep logs, tracked by: total count of sleep a day, total minutes, total time in bed.
weightLogInfo_merged	CSV	Weight track by day in kg and pounds over 30 days. Including the calculation of BMI for eight users.

Note: The archive consists of 18 CSV documents. Each CSV document in the Fitbit dataset represents a different aspect of the quantitative data tracked by the Fitbit device, such as steps, distance, calories burned, and sleep. Each row in the CSV file represents a single time point for a specific user, and each user has a unique ID. Since data is tracked by day and time, each user will have multiple rows in each CSV document.

To gain insights into the usage behaviors of the Fitbit dataset, we started by analyzing high-level trends. A summary of each variable in the dataset is shown in Table 4, revealing missing data across all categories. The daily activity category includes steps, distance, calories, and minutes spent in sedentary, lightly active, fairly active, and very active activity levels. On average, users take 7638 steps per day, with a median of 7406. The maximum and minimum number of steps recorded in a day are 36019 and 0, respectively. This suggests the presence of missing data, as it is unlikely that a user would stay still for an entire day. Users cover an average distance of 5.49 km per day, with a median of 5.245 km. The mean number of calories burned per day is 2304, with a median of 2134. The majority of participants are lightly active, spending an average of 991 minutes or 16 hours in sedentary activities. In the daily intensity category, the dataset includes minutes and distance for sedentary, lightly active, fairly active, and very active activity levels. Users spend an average of 991.2 minutes per day in sedentary activities, with a median of 1057.5 minutes. The mean minutes of lightly active activity per day is 192.8, with a median of 199. Users cover an average distance of 3.341 km in lightly active activities per day, with a median of 3.365 km. For fairly active activities, the mean time spent per day is 13.56 minutes, with a median of 6 minutes, and users cover an average distance of 0.56 km per day, with a median of 0.24 km. The mean time spent in very active activities per day is 21.16 minutes, with a median of 4 minutes, and users cover an average distance of 1.53 km per day, with a median of 0.21 km. In the sleep category, the dataset includes minutes of sleep and the number of awakenings per night. On average, users spend 419.5 minutes asleep, with a median of 433 minutes. The mean number of times users are in bed per night is 458.6, with a median of 463.0.

In order to gain a deeper understanding of the dataset, we analyzed the hourly data over a 24-hour period. We investigated the total average calories consumed, the intensity expended, and the number of steps taken per hour. Figure 2 illustrates the average total calories consumed per hour. In the context of Fitbit, calories refer to the amount of energy expended by an individual throughout the day, including various activities like exercise, daily tasks, and even basic body functions such as digestion and breathing. Fitbit calculates the total number of calories burned by taking into account different factors, such as a person's basal metabolic rate, activity level, and the number of steps taken. Figure 3 displays the average total intensity expended per hour. Intensity refers to the level of physical activity or exercise performed by the user. Fitbit devices track different intensity

Table 4: The table illustrates the overview and summary of the Fitbit dataset.

Category	Variables Name	Min.	1st Qu.	Median	Mean	3rd Qu.	Max	Instances	N
Daily-level variables	Calories	0.0	1828	2134	2304	2793	4900	940	33
	TotalSteps	0.0	3790	7406	7638	10727	36019	940	33
	TotalDistance	0.0	2.620	5.245	5.490	7.713	28.030	940	33
	SedentaryMinutes	0.0	729.8	1057.5	991.2	1229.5	1440.0	940	33
	LightlyActiveMinutes	0.0	127.0	199.0	192.8	264.0	518.0	940	33
	FairlyActiveMinutes	0.0	0.0	6.0	13.56	19.0	143.0	940	33
	VeryActiveMinutes	0.0	0.0	4.0	21.16	32.00	210.0	940	33
	SedentaryActiveDistance	0.0	0.0	0.0	0.001606	0.00	0.11	940	33
	LightActiveDistance	0.0	1.945	3.365	3.341	4.782	10.710	940	33
	ModeratelyActiveDistance	0.0	0.0	0.24	0.5675	0.80	6.48	940	33
	VeryActiveDistance	0.0	0.0	0.210	1.503	2.053	21.920	940	33
	LoggedActivitiesDistance	0.0	0.0	0.0	0.1082	0.0	4.9421	940	33
	TotalSleepRecords	1.0	1.0	1.0	1.12	1.0	3.0	410	24
	TotalMinutesAsleep	58.0	361.0	433.0	419.5	490.0	796.0	410	24
	TotalTimeInBed	61.0	403.0	463.0	458.6	526.0	961.0	410	24
Hourly-level variables	Calories	42.0	63.0	83.0	97.39	108.0	948.0	22099	33
	TotalIntensity	0.0	0.0	3.0	12.04	16.0	180.0	22099	33
	AverageIntensity	0.0	0.0	0.05	0.2006	0.2667	3.0	22099	33
	StepTotal	0.0	0.0	40.0	320.2	357.0	10554.0	22099	33
Minute-level variables	Calories	0.0	0.9357	1.2176	1.6231	1.4327	19.7499	1325580	33
	Intensity	0.0	0.0	0.0	0.2006	0.0	3.0	1325580	33
	METs	0.0	10.0	10.0	14.69	11.00	157.0	1325580	33
	Sleep_value	1.0	1.0	1.0	1.096	1.0	3.0	187978	24
	Steps	0.0	10.0	0.0	5.336	0.0	220.0	1325580	33

Note: The dataset is organized into three categories of aggregated data: daily level, hourly level, and minute level variable. For each numerical variable, we have listed the variable name, minimum value, 1st quartile, median, mean, 3rd quartile, maximum value, number of instances, and number of unique users. This table was adopted from [15, 7, 59].

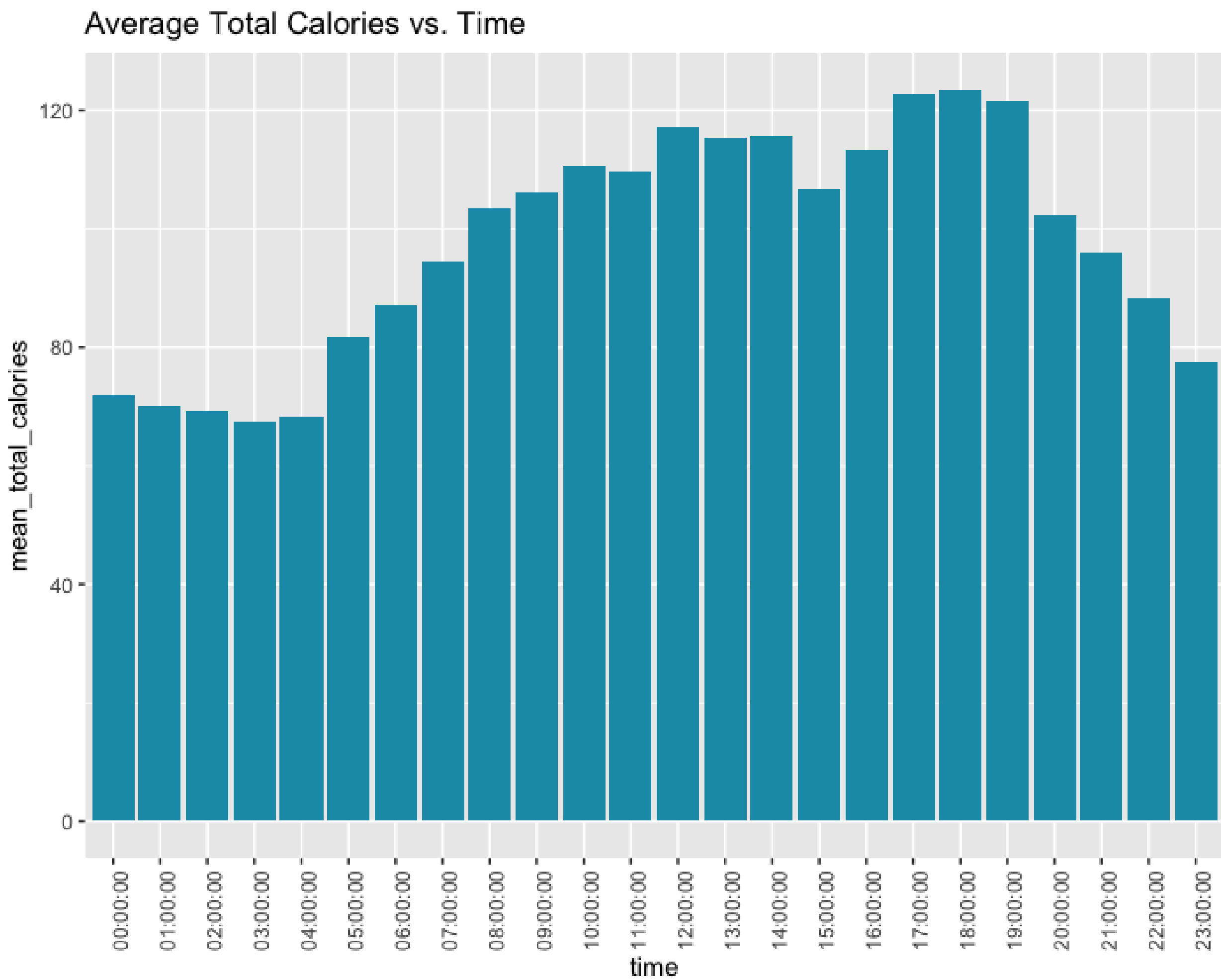
levels, such as sedentary, lightly active, fairly active, and very active, based on the number of steps taken, distance covered, and the duration of activity. By using these intensity levels, users can gain insights into how much physical activity they engage in throughout the day, and make informed decisions about their fitness goals and lifestyle. Figure 4 indicates the average total steps per hour. Steps represent the number of steps taken by the user while wearing their Fitbit device. Fitbit uses accelerometers to monitor the movement of the user's wrist, and this data is then used to estimate the number of steps taken. The overall trend observed in Figure 2, 3, and 4 reflects the consistent behavior of humans where they tend to be more active during the daytime and less active during the night time. Furthermore, the trends observed in these figures are consistent with each other, which further confirms the accuracy of the dataset.

To gain a deeper understanding of the distribution of the user type, we classified users into different activity levels – sedentary, low active, somewhat active, active, and highly active – based on the classification scheme proposed in [10]. These categories were determined by the average number of daily steps taken by each user. Specifically, the sedentary group includes users who take less than 5,000 steps per day, the lightly active group includes users who take between 5,000 and 7,499 steps per day, the somewhat active group includes users who take between 7,500 and 9,999 steps per day, the fairly active group includes users who take between 10,000 and 12,499 steps per day, and the highly active group includes users who take more than 12,500 steps per day. As shown in Figure 5, in this dataset, 24.2% of users are sedentary, 27.3% are lightly active, 27.3% are somewhat active, 15.2% are fairly active, and 6.1% are highly active.

3.3 General Observations of Dataset

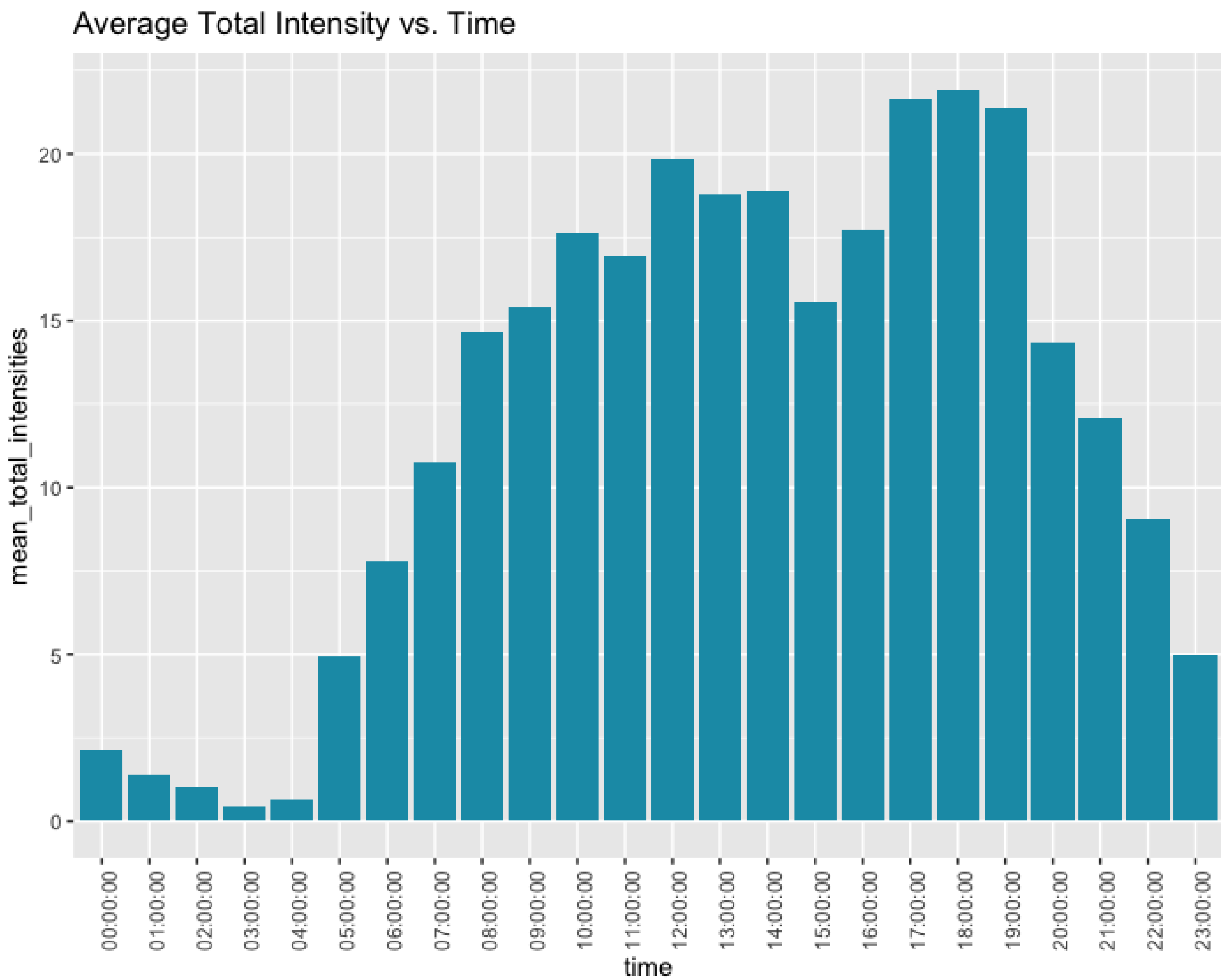
To better understand the distribution of each variable, we conducted a detailed analysis of each variable. Table 5 summarizes this analysis, which involved identifying and removing missing data for each variable, recalculating key statistics (minimum, 1st quartile, median, mean, 3rd quartile, maximum), and also counting the number of instances, the percentage of missing data, and the number of unique users. Our analysis revealed several limitations in identifying missing data in such behavioral datasets, as well as some trends that are common to these types of datasets. For

Figure 2: Average total calories consumed.



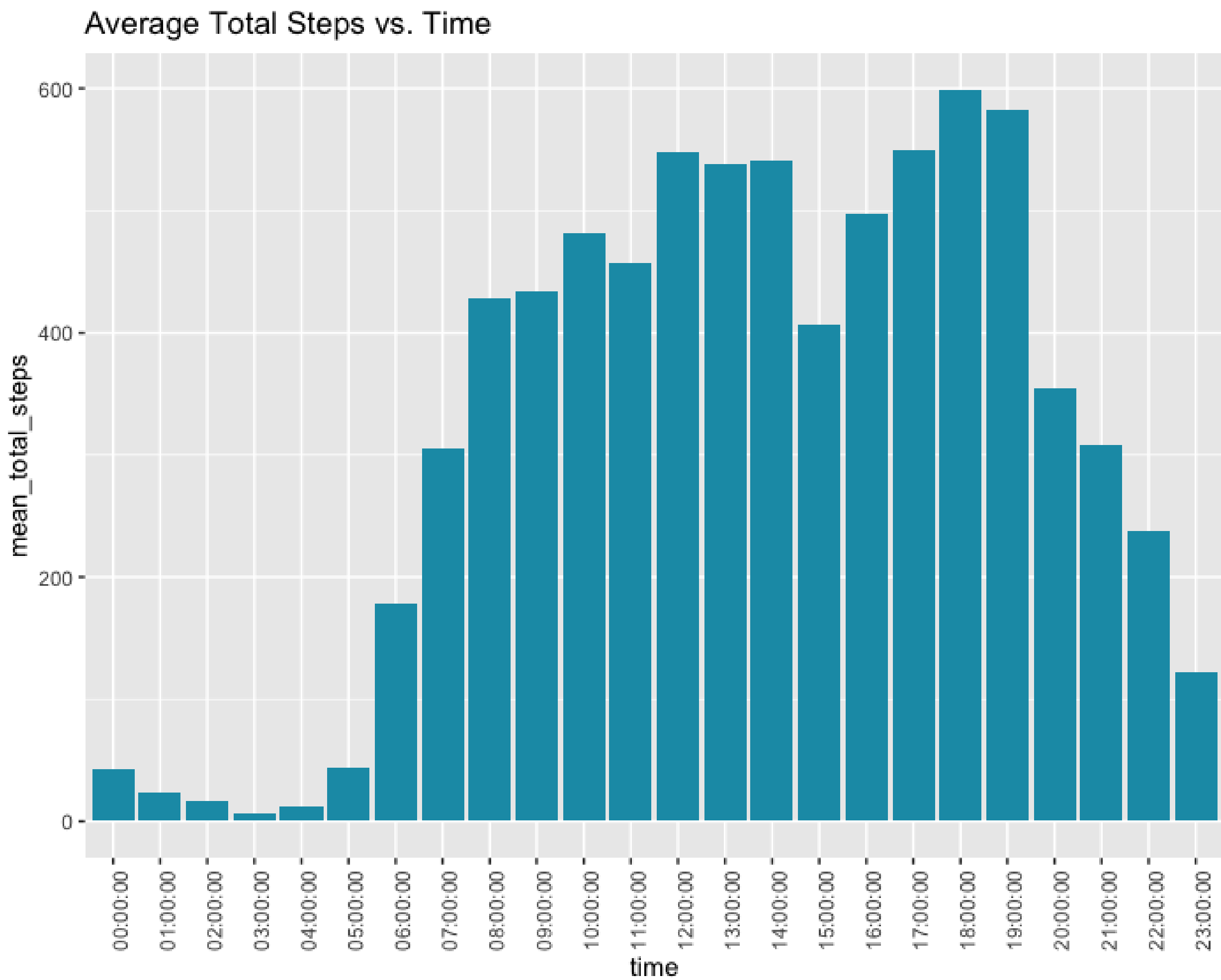
Note: This figure is adapted from [15], illustrating the average total calories consumed per hour during a 24-hour window.

Figure 3: Average total intensity expended.



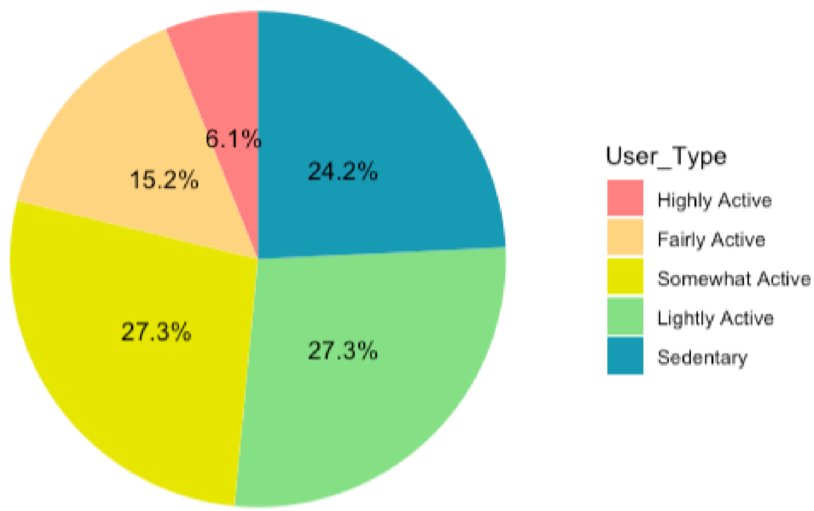
Note: This figure, which illustrates the average total intensity expended every hour during a 24-hour window, is adopted from [15].

Figure 4: Average total steps taken.



Note: The figure depicting the average total steps taken in each hour during a 24-hour window is adapted from [15].

Figure 5: User type distribution.



Note: The presented figure illustrates the distribution of user types based on their steps taken, which is adapted from [59].

instance, with the variable "calories," we know that Fitbit or other smartwatches calculate this variable by tracking energy burns every minute, hour, or day. Therefore, if the value equals zero, we can confidently say that there is missing data. However, for variables such as "steps" and "distance," it is harder to determine missing data since it is possible for someone to be sedentary or stationary for a period of time, such as sitting for an hour without taking any steps. For example, Table 5 shows a high percentage of missing data for "SedentaryActiveDistance," "Intensigy," and "Steps." As a result, missing data in these behavior datasets can be more complex, and the missingness can be hard to classify, which can affect subsequent analyses.

3.4 Methods

3.4.1 Notation

Based on the definitions provided by Wu et al. [105] and Lin et al. [65] for time series data, the data gathered from wearables are specific to each user. Therefore, we use the notation $U = u_1, ..., u_i, ..., u_I$ to represent the participants in the dataset, where I is the number of participants. The collection dates are described as a vector $t_i = t_i^1, ..., t_i^j, ..., t_i^{J_i}$ in chronological order, where t_i^j denotes the date collected from participant u_i on the j-th day, and J_i is the total number of days collected. The time series recorded on t_i^j is denoted by $x_i^j = x_i^{(j,1)}, ..., x_i^{(j,\tau)}, ..., x_i^{(j,T)}(x_i^j \in \mathbb{R}^T)$. Each measurement $x_i^{(j,k)}$ is associated with a timestamp.

3.4.2 Temporal Properties (TP)

Each instance collected in a wearable sensory device, such as Fitbit, is associated with a timestamp that allows us to order them chronologically and generate sequential properties such as trends and frequencies. Sequential properties of user i are defined as follows:

$TP_i = (x_i^{j,T} - x_i^{(j,T-1)}), ..., (x_i^{j,T-k} - x_i^{(j,T-k-1)}), ..., (x_i^{j,T-\sigma} - x_i^{(j,T-\sigma-1)})$

Where σ denotes the time gap used to calculate the difference between two consecutive days.

Table 5: Overview and summary of the Fitbit dataset after removing missing data (0 values).

Category	Variables Name	Min.	1st Qu.	Median	Mean	3rd Qu.	Max	Instances	Missing	N
Daily-level variables	Calories	52	1834	2144	2313	2794	4900	936	0.43%	33
	TotalSteps	4	4923	8053	8319	11092	36019	863	8.19%	33
	TotalDistance	0.01	3.37	5.59	5.99	7.91	28.03	862	8.30%	33
	SedentaryMinutes	2.0	730.0	1058.0	992.3	1230.0	1440.0	939	0.11%	33
	LightlyActiveMinutes	1.0	148.0	209.5	211.7	272.0	518.0	856	8.90%	33
	FairlyActiveMinutes	1.0	8.0	16.0	22.9	30.0	143.0	556	40.86%	33
	VeryActiveMinutes	1.0	10.0	27.0	37.5	55.00	210.0	531	43.51%	33
	SedentaryActiveDistance	0.01	0.01	0.01	0.02	0.02	0.11	82	91.28%	33
	LightActiveDistance	0.01	2.37	3.59	3.67	4.91	10.71	855	9.04%	33
	ModeratelyActiveDistance	0.01	0.34	0.66	0.96	1.20	6.48	554	41.06%	33
	VeryActiveDistance	0.02	0.60	1.76	2.68	3.53	21.92	527	43.94%	33
	LoggedActivitiesDistance	1.96	2.09	2.25	3.18	4.86	4.94	32	96.60%	33
	TotalSleepRecords	1.0	1.0	1.0	1.12	1.0	3.0	410	0.00%	24
	TotalMinutesAsleep	58.0	361.0	433.0	419.5	490.0	796.0	410	0.00%	24
	TotalTimeInBed	61.0	403.0	463.0	458.6	526.0	961.0	410	0.00%	24
Hourly-level variables	Calories	42.0	63.0	83.0	97.39	108.0	948.0	22099	0.00%	33
	TotalIntensity	1.0	6.0	13.0	20.46	25.0	180.0	13002	41.16%	33
	AverageIntensity	0.02	0.1	0.21	0.34	0.41	3.0	13002	41.16%	33
	StepTotal	1.0	103.0	287.0	552.7	643.0	10554.0	12802	42.07%	33
Minute-level variables	Calories	0.70	0.94	1.22	1.62	1.43	19.75	1325573	0.00%	33
	Intensity	1.0	1.0	1.0	1.25	1.0	3.0	213478	83.90%	33
	METs	6.0	10.0	10.0	14.69	11.00	157.0	1325573	0.00%	33
	Sleep_value	1.0	1.0	1.0	1.10	1.0	3.0	187978	0.00%	24
	Steps	1.0	11.0	230.0	35.38	48.0	220.0	199910	84.92%	33

Note: The dataset is organized into three categories of aggregated data: daily level, hourly level, and minute level variable. For each numerical variable, we have listed the variable name, minimum value, 1st quartile, median, mean, 3rd quartile, maximum value, number of instances, percentage of missing data, and number of unique users.

3.4.3 Contextual Information (CI)

We also integrate the previous instance collected for the same participant in our model to leverage contextual information surrounding missing intervals. Assuming that the available data set from earlier dates for the target missing instance may include additional features to aid in imputing the missing part, we define the contextual information for user i as follows:

$CI_i = (x_i^{(j,T-1)}), ..., (x_i^{(j,T-k-1)}), ..., (x_i^{(j,T-\sigma-1)})$.

3.4.4 Linear & Polynomial Regression (LiR)

We applied ordinary least squares (OLS) regression analysis to develop linear models for imputing missing values in the Fitbit dataset, using the temporal and contextual properties as predictors. OLS regression analysis is a statistical technique used to determine the relationship between a dependent variable and one or more independent variables. OLS aims to identify the line (or hyperplane in higher dimensions) that minimizes the sum of the squared differences between the predicted and actual values. The resulting string is then used to make predictions or estimate values of the dependent variable based on the independent variable values. OLS is widely used in machine learning and statistical analysis.

3.4.5 Logistic Regression (LoR)

Our motivation for implementing this model was based on its ability to handle dense and sparse input data. In addition, the regression model we used incorporates L2 regularization, also known as Ridge Regression, a standard machine-learning technique to prevent overfitting. L2 regularization involves adding a penalty term to the cost function during training that discourages the model from relying too much on any one feature, effectively shrinking the coefficients toward zero. The penalty term is proportional to the squared magnitude of the coefficients, hence the name L2 (since it involves the L2 norm of the coefficients).

In our model, the loss function is the least-squares error (LSE), commonly used in regression analysis to measure the difference between predicted and actual values. In linear regression, the goal is to find a line that best fits the data by minimizing the sum of the squared differences between the predicted values and the actual values. The LSE is the sum of the squared differences between the predicted and actual values divided by the number of data points. By incorporating L2 regularization and using the LSE loss function, our model can effectively handle overfitting and produce accurate predictions for missing data.

$$Loss = LSE(y_i, \widehat{y_i}) + \lambda \sum_{i=1}^{N} w_i^2 \tag{1}$$

Where y_i is the actual value, $\widehat{y_i}$ is the predicted value, and n is the number of data points. The LSE is commonly used as a loss function because it is easy to compute, differentiable, and has desirable mathematical properties that make it a good fit for linear regression.

This model tries to minimize the error between the actual y and the predicted $\widehat{y}$ by adding the regularisation term as an optimization during the process. The L2 regularization is proportional to the square of the L2 norm of the model's weights and added to the optimized loss function. By adding this term, the model is penalized for using large weights and is encouraged to use smaller weights instead, which can help to prevent overfitting and improve the generalization performance of the model. The optimal weights are found by minimizing the sum of the loss function and the regularization term, typically done using an optimization algorithm such as gradient descent.

3.4.6 Decision Tree (DT)

The decision tree model builds the estimation model as a tree structure. It first decomposes an extensive dataset into smaller and smaller subsets while, at the same time, an associated decision tree is developed incrementally. After the training, a tree with the decision and leaf nodes will be formed. A decision node has two or more branches, each representing values for the attribute tested. The leaf node represents an estimation of the numerical target—the topmost decision node in a tree corresponding to the root node's best predictor. Decision trees can handle both categorical and numerical data. We used the mean squared error (RMSE) as a feature selection criterion for the supported standards; we used the average of each terminal node's value to minimize the L2 loss and applied different variances for turning the tree.

3.4.7 Random Forest (RF)

A random forest estimator is a machine learning algorithm that uses an ensemble of multiple decision trees trained on different sub-samples of the dataset. Each tree in the random forest is trained perfectly on its sample data. In regression problems, the output is obtained by averaging the results of all the trees, which improves the predictive accuracy and helps to control overfitting. Random forest is based on Bootstrap and Aggregation (or bagging), which involves generating multiple bootstrap samples of the original data and training decision trees on each piece. The aggregated result from all the trees in the forest is then used as the final output. The random forest can be used for both regression and classification problems. It is a popular machine-learning algorithm due to its high accuracy and ability to handle large datasets with high-dimensional features.

We implemented regression models and used the extracted TP and CI properties as predictors to calculate the missing values. We followed a two-stage procedure to select the appropriate model for our study. In the first stage, we used the original dataset with missing values deleted, resulting in a complete-case dataset denoted as X_i'. In the second stage, we incorporated the TP and CI properties extracted from the original dataset and characterized the resulting dataset as $X_i'' = X_i' \times TP_i \times CI_i$. We referred to these datasets as FD and FDTP, respectively. Finally, we constructed regression-based prediction models on both datasets using OLS regression analysis or Ordinary Least Squares regression analysis. It is a statistical technique to estimate the relationship between a dependent variable and one or more independent variables. OLS aims to find the line (or hyperplane in higher dimensions) that minimizes the sum of the squared differences between the predicted and actual values. The resulting string is then used to make predictions or estimate the dependent variable's values based on the independent variables' values. We denoted the prediction models as $*$_FD and $*$_FDTP, where $*$ can be replaced with LiR, LoR, DT, and RF.

Table 6: The table presented provides descriptions for each model we implemented.

Models	Description
	Regression Models
LiR	OLS-based linear regression model
LoR	Ridge Regression model
DT_i	Decision tree models, i stands for different tuning strategies
RF_i	Random Forest models, i stands for different tuning strategies
	Data Properties
FD	Temporal closeness instance + contextual information
FD_{DT}	FD + daily trend sequence

3.5 Experimental Results

In this section, we performed comprehensive experiments aimed at addressing the following research questions:

- **RQ1**: To what extent does the proposed method accurately estimate our missing data?
- **RQ2**: How does the inclusion of TP and CI as predictors compare to a model without TP and CI?
- **RQ3**: How do different tuning policies impact model performance?
- **RQ4**: Which model is best suited for visualizing the missing data, which is utilized in the subsequent stages of our research?

3.5.1 Experiment setting

In our evaluation, we conducted experiments on the Fitbit data discussed in Section 3.2. The dataset was divided into two subsets: an 80% training set and a 20% test set. We assessed the dataset using four models, namely LiR, LoR, DT, and RF, as detailed in Table 6.

3.5.2 Evaluation protocol

We employed the following metrics to evaluate the performance of the effort prediction model. First, the overall fitness of the model was assessed by calculating the Adjusted R^2. Adjusted R-squared penalizes excessive use of variables unrelated to the dependent variable, and it is useful when comparing the performance of models with different numbers of predictors. It is calculated as follows:

$$R^2_{adj} = 1 - [\frac{(1 - R^2)(n - 1)}{n - k - 1}] \tag{2}$$

Where R^2 is the standard R-squared value, n is the sample size, and k is the number of independent variables in the model. The adjusted R-squared value ranges from 0 to 1, with higher values indicating a better fit of the model to the data.

The Root Mean Square Error (RMSE) is a standard metric used to evaluate the accuracy of a regression model. It measures the difference between the predicted and actual values in a dataset. The RMSE is calculated by taking the square root of the average of the squared differences between the predicted values and actual values. The formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{z}\sum_{i}(v_i - \hat{v}_i)^2} \tag{3}$$

Here, $\hat{v}$ and v denote the predicted and actual values, respectively, and z represents the total number of predicted values.

3.5.3 Performance

In our experiments, we assessed the performance of various methods under different configurations of trend sequence and contextual information using the Fitbit dataset to predict TotalSteps. The results from Table 7 yielded several key observations:

i. The *LiR* model consistently outperforms all other methods across various conditions. For instance, *LiR* demonstrates improvements in RMSE scores ranging from 11.6% to 90.7% for temporal closeness instances, from 19.3% to 93.2% for temporal closeness and trend sequence, and from 11.7% to 90.4% for temporal closeness and contextual information.

Table 7: The table presented reports the $RMSE$ and adjusted R^2 for each model.

Models	***RMSE***	**Adjusted** R^2
LiR_{FD}	69.68	0.969
LoR_{FD}	753.98	0.620
DT_{FD}	102.18	0.935
RF_{FD}	79.21	0.961
$LiR_{FD_{DT_1}}$	**66.27**	0.972
$LoR_{FD_{DT_1}}$	983.27	-1.675
$DT_{FD_{DT_1}}$	126.39	0.900
$RF_{FD_{DT_1}}$	82.13	0.957
$LiR_{FD_{DT_2}}$	75.52	0.965
$LoR_{FD_{DT_2}}$	790.93	0.91
$DT_{FD_{DT_2}}$	109.83	0.924
$RF_{FD_{DT_2}}$	85.58	0.958

ii. Among all the compared methods, the model that consistently outperforms the others is $LiR_{FD_{DT_1}}$, which applies temporal closeness and daily trend sequence to the Fitbit dataset.

iii. The linear model consistently delivers superior performance across all conditions, while the logistic model consistently performs the worst. This suggests that variations of logistic models can be excluded from future analyses, and the linear model is the most suitable choice for subsequent research. This answers our research question RQ1 that the proposed method can estimate the missing data with a good confidence

In addition, we also conducted experiments involving different training and testing sets while maintaining a fixed gap size for calculating additional data properties. The gap size used in this experiment was based on the daily intervals used in the Fitbit and other wearable applications. The result (shown in Table 7) indicates that the *Lir* model consistently outperforms other compared methods across various gap sizes. Notably, the smaller the time gap used when generating trend sequences (e.g., FD_{DT_1} at 24 hours or FD_{DT_1}) at 48 hours), the better the performance of the *LiR* and *RF* models. However, this trend does not hold for the *LoR* and *DT* models.

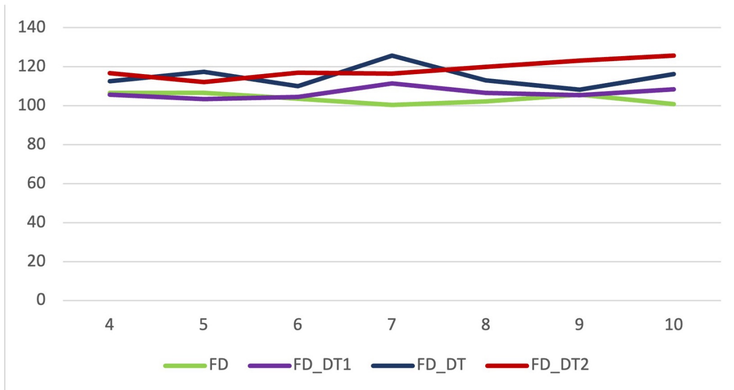
We also conducted experiments to explore the impact of different parameters, specifically the depth of the decision tree. All other parameters were kept at their default settings. Fig. 6 displays the RMSE of the tuning parameter while controlling for others. The result indicates that the tuning strategy has a relatively low influence on the performance of the DT model for our dataset. This answers our research question RQ3.

To answer RQ4, which model is best suited for visualizing the missing data, we assessed the Goodness-of-Fit by reporting the adjusted R^2 value for each model. The result revealed that the adjusted R^2 values are fairly similar across the models. Thus, theoretically, any of these three models can be used for future research. However, the LoR model, with a close to zero R^2 value, should be excluded from consideration.

3.6 Discussion

Throughout the process of developing and testing models for automatic missing data estimation, we encountered significant limitations in relying solely on an algorithmic approach. Identi-

Figure 6: Effects of the tuning strategy.



Note: The figure presented illustrates the effects of the tuning strategy on the DT model in the algorithmic approach exploration.

fying missing data posed several challenges. Firstly, as outlined in Table 2, individuals use their wearable devices for a range of activities, including walking, running, biking, and hiking. These activities are often sporadic and varied, making it difficult for an algorithm to distinguish missing activities without confirmation from the user.

Furthermore, even in cases of daily step counts, which might seem straightforward to identify when data is missing, there is a notable issue. Typically, a user is expected to move during the day, ensuring that there are no days with zero daily steps. While algorithmic estimation can be employed for such data and yield high accuracy, it introduces another complication: the time gap. Estimating daily step counts may be accurate, but determining whether the hourly step data is missing or if the user is simply in a sedentary state can be challenging.

These limitations highlight the need to transition from an algorithmic approach to a human-in-the-loop approach. This shift allows us to harness the strengths of automated processes while benefiting from human judgment, common sense, and contextual understanding. It is evident that automated systems alone may not effectively address the challenges of identifying and estimating missing data, and human expertise is essential to ensure the quality and reliability of the estimation process.

4.0 Towards Understanding Barriers: Missing Data in Personal Informatics Systems

As discussed in Section 3, the purely algorithmic approach to handling missing data has found it challenging to identify individual differences using general models. These limitations have prompted us to adopt a human-in-the-loop approach to address the challenges posed by missing data in self-tracking. In this section, we have undertaken semi-structured interviews to gain insights into the specific requirements of individuals who utilize wearable devices for tracking purposes.

4.1 Method

4.1.1 Research Design

This study aimed to explore the influence of missing data on the usage, utility, and associated behaviors of Personal Informatics (PI) tools. It focused on addressing the following research questions:

- What are users' expectations regarding the consistency and completeness of PI data?
- How does the presence of missing PI data affect these expectations?
- How can the design of PI tools be enhanced to assist users in recognizing and addressing missing data within their tracking activities?

To address our research inquiries, we conducted semi-structured interviews. Eligible participants were required to possess or have regular access to Personal Informatics (PI) systems, like Fitbit, iWatch, Garmin, Strava, etc., for a minimum of three months. They should also have engaged with their systems' dashboards and reports for self-reflection at least once a month. Our study involved twenty individuals. Additionally, we administered a concise survey using a Likert scale to gather demographic details and assess users' perceptions of the impact of missing data on their PI objectives and their overall attitude toward incomplete PI data.

4.1.2 Participants

We initiated our recruitment process by sending recruitment emails to private training groups, including Panther Cycling Club and Club Triathlon of Pittsburgh. Additionally, we distributed the recruitment questionnaire through the crowdsourcing website Prolific to identify individuals who regularly use Personal Informatics (PI) tools. A total of 314 individuals responded to the questionnaire. We then applied the following inclusion criteria: 1) had been using PI tools for over three months, 2) engaged in data reflection at least once a month, and 3) collected data to achieve specific goals. We randomly selected participants from the three recruitment sources. As a result, our study included 20 participants (11 female, 9 male) ranging in age from 19 to 56 years, with a mean age of 32.875 and a standard deviation of 9.34. All participants resided in the United States. The number of participants was determined based on Qualitative Research and Saturation Criteria [87], and we concluded the surveys when no new information emerged in subsequent interviews. As indicated in Table 8, all participants possessed substantial experience with tracking.

4.1.3 Procedure

The design of this study received approval from our University's institutional review board (IRB). Individuals who completed the pre-screening Qualtrics survey were subsequently contacted via email to arrange virtual interviews, which lasted up to an hour. Notably, this research was carried out during the COVID-19 pandemic, necessitating the use of Zoom for conducting interviews. The interviews were recorded for analysis. During the interviews, participants were initially provided with a comprehensive explanation of the study and requested to give their informed consent. Subsequently, participants were asked a series of questions to elucidate the Personal Informatics (PI) tools they employed, their underlying motivations, and their usage patterns. They were also requested to demonstrate how they engaged in data reflection and how instances of missing data occurred and impacted this process. Participants were then prompted to express their attitudes toward missing data through descriptive sentences and a 5-point Likert scale. To conclude, participants were invited to provide suggestions for mitigating the effects of missing data. The sequence of the user study is shown in Figure 7. We compensated participants with a payment of $15 for their valuable contributions to the study.

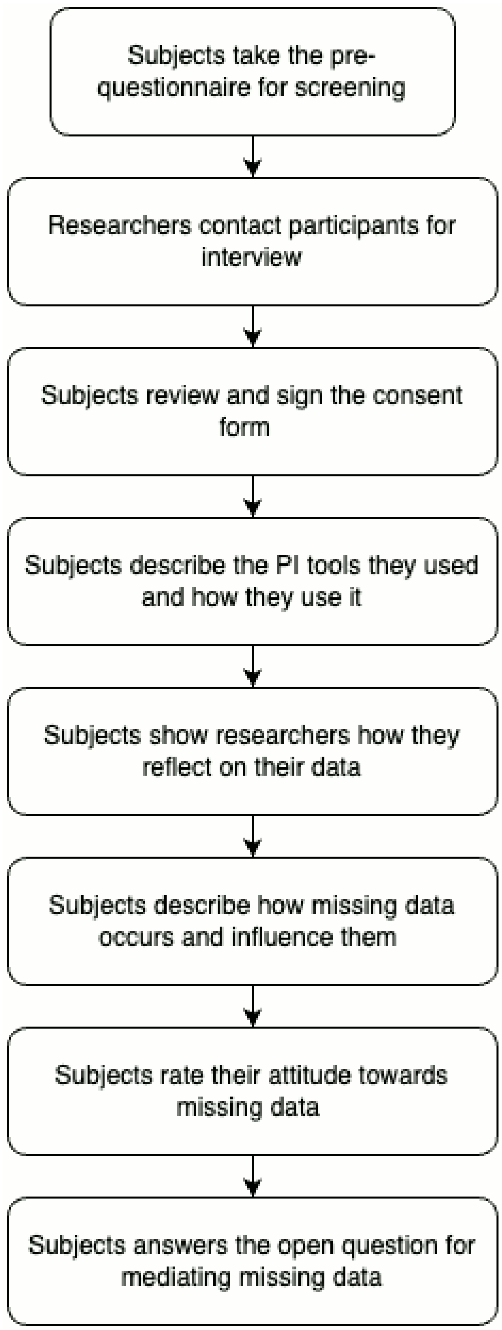


Figure 7: Flowchart that depicts the sequence of the interview study.

4.1.4 Data Preparation and Analysis

The audio files from the interview were imported and machine transcribed using the cloud-based platform Atlas.ti. Subsequently, two researchers, both of whom are authors of this paper, conducted open coding to discern prevalent themes and patterns in the participants' responses. They further explored connections among these codes in a process known as axial coding, as described by Scott ([88]). Following this, the first researcher and a third researcher, who is not an author of this paper, individually examined and coded 10% of the transcripts, utilizing the previously identified themes. Their initial level of agreement (inter-rater reliability), expressed as a percentage, was 0.769, which surpassed the expected chance agreement of 0.56, as proposed by Krippendorff ([45], p. 224-226). The first and third researchers engaged in discussions to address any discrepancies and update the existing codebook. Subsequently, the first researcher and a fourth researcher who is not affiliated with this paper completed the coding of the remaining transcripts. The inter-rater reliability, assessed using Krippendorff's alpha, was determined to be 0.73, with an agreement level of 0.99.

4.2 Results

Our study found notable user behaviors and perceptions in contexts where personal informatics tools have missing data. These findings revolve around the challenges associated with data capture and goal tracking, leading to the identification of two distinct user categories: ***trainees*** and ***maintainers***. Trainees are individuals with specific, often involving professional athletic training who utilize data for informed decision-making aimed at enhancing performance-related metrics (e.g., cardiac efficiency or interval times). In contrast, maintainers have more general health improvement and maintenance objectives and employ data to monitor milestones such as daily step counts and active exercise hours. Given these markedly different goals, we present our results to provide insights from the perspective of *both* user classes.

Table 8: Participants' demographics, tracking background, and frequency of reflection

PID	Age range	Gender	Occupation	Wearables	Duration	Category	Main motive of use	Frequency and types of reflection
P1	32-43	Male	Cycling coach	Garmin watch	13 years	Trainees	Training performance, analysis for cycling data, balance training load.	Reflect on physiological data after biking, deeper analysis on the weekend to check performance, modify training load, and analyze performance per season and year.
P2	19-31	Female	Service	Apple watch	7 years	Maintainers	Maintain heart rate, maintaining weight, be active.	Check data after workouts, check meal data daily on the app for nutrition and reflect on weekends for meal summary.
P3	32-43	Male	Teacher	Garmin Phoenix five	13 years	Trainees	Replicate best performance for the race, inform training schedule	Reflect on data after running, reflect on weekly data to analyze and identify best performance based on road conditions, pace, and duration.
P4	32-43	Male	Software engineer	Kronos watch	1 year	Maintainers	Maintaining weight, improve health	Check to see if hit daily targets, check weekly data to see trends.
P5	44-56	Female	Unemployee	VeryFit watch	2 years	Maintainers	Maintaining health, losing weight	Reflect on weekly data and daily data to see if hit targets.
P6	19-31	Female	Desk job	Apple watch	> 6 years	Maintainers	Tracking activities, checking calories burned.	Multiple times a day to see if they hit daily targets, reflect three times a month to see summaries reflect, weekly to see if they hit weekly targets.
P7	19-31	Male	Student	Suunto watch	3 years	Trainees	Tracking and analyzing cycling data, log exercises.	Multiple times a day, reflect after cycling for performance
P8	19-31	Female	Student	Garmin watch	2 years six mo	Maintainers	Tracking activities, hit exercise target.	Check during workouts, reflect at least once a day to see if hit targets.
P9	32-43	Female	Unemployee	Fitbit	> 10 years	Maintainers	Tracking activities, log workouts.	Check data after workouts, check weekly to see exercise types and duration.
P10	44-56	Male	Learning Consultant	Garmin watch	5 years	Trainees	Maintain certain pace during marathon	Reflect after running for pace, heart rate and road condition
P11	19-31	Female	Desk job	Garmin watch	5 years	Trainees	Training to advance calisthenics	Reflect three times a week for how many times reached the goal
P12	32-43	Female	Desk Job	Fitbit	5 years	Maintainers	Tracking exercises, maintaining weight, being active	Check multiple times a day during exercise, reflect once a day to see if hit targets.
P13	19-31	Female	Unemployee	Fitbit	1 year	Maintainers	Tracking activities, be active	Checking data multiple times a day and during exercise, reflect daily to see if hit targets.
P14	19-31	Female	Student	Apple Watch, Whoop band	4 years	Trainees	Training for Brazilian Jiu-Jitsu competition, improve performance	Reflect daily to check recovery score, inform the types of training the body is ready for the day, reflect weekly and monthly to check trends/patterns of the week&month.
P15	19-31	Female	Student	Apple Watch	2.5 years	Maintainers	Tracking exercises, losing weight	Reflect daily to check if hit targets, check during exercise to see progress.
P16	19-31	Female	Teacher	Apple Watch	2 years	Trainees	Training for weight lifting, building muscles	Checking after training for performance: duration, weight; reflect to inform training load
P17	32-43	Male	Pilot	Whoop band	1.5 years	Trainees	Training for hike Mount Rainier	Reflect after hiking for speed, heart rate, and altitude of the mountain.
P18	44-56	Male	Maintenance technician	Garmin Vivoactive Three	8 years	Trainees	Training for running	Checking data for time and heart rate during training. Balance training load. Reflect after running for hill work, pace, distance, elevation, and duration.
P19	32-43	Male	Teacher	Misfit Vapor X	4 years	Trainees	Training for cycling, making sure to hit endurance targets	Reflect weekly to inform training in the following week, reflect after cycling to check for performance
P20	19-31	Male	Wine tasting host	Apple Watch	> 1 year	Maintainers	Tracking workout, stay active	Checking data multiple times per day for progress, reflect weekly for summaries

4.2.1 Usage of the PI Tools

To provide context, we begin by describing how trainees and maintainers incorporated PI tools into their daily routines. The 20 participants interviewed made use of a wide range of PI tools. Table 9 summarizes these tools, separating them between the two classes of users. We conducted an analysis of each tool, evaluating their functionality across four dimensions:

- Data export: The ability to utilize data collected by the tool in other applications.
- Data visualization: The provision of visual representations of activity and progress toward predefined goals.
- Data analysis: The capability to identify performance trends and adjust future objectives.
- Social features: The inclusion of options for sharing physical activities or achievements on social media.

The analysis of feature presence involved considering participants' explicit descriptions of use, as well as our evaluation of manufacturer marketing materials and technical manuals. For application-based tools, we also conducted independent installations and explorations. The results of this analysis are reported in the last four columns of Table 9. In the table, a circle indicates that we found the feature dimension to be present in the tool, while a dash indicates that, if assessed, the feature was not present. A solid shading in the table signifies that at least one participant reported using the feature. An empty circle suggests that the feature was present in the tool but was not used by any participants in the interviews.

This analysis revealed several intriguing patterns. One of the most prominent findings is that many of the *same* tools were utilized in distinct ways by different classes of users. In particular, ***trainees*** showed a strong preference for using data export and data analysis functionalities over built-in (or even non-present) data visualization features. Conversely, ***maintainers*** displayed a different usage pattern, favoring visualization features. The sentiments and motivations behind these contrasting behaviors were captured during the interviews.

Trainees often expressed specific expectations and needs for data capture and analysis. For example, P3, a ***trainees***, mentioned the importance of collecting a specific data pattern: *"I had a foot pod, which would measure my cadence when I was running."* Similarly, P1, also a ***trainees***, believed that *"[Golden Cheetah] is a much more powerful tool to analyze performance"* compared

Table 9: PI wearables/applications used by participants.

Category	**PI tools used**	**Usage of the PI tools**			
		Data export	Data visualization	Data analysis	Social features
Trainees	**Apple Watch**	●	○	●	○
	Bike Computers	●	○	—	—
	Couch to 5k (app)	●	—	—	—
	Garmin Fenix Watch	●	○	●	○
	Garmin Vivoactive Watch	●	●	●	○
	Garmin Watch	●	○	●	○
	Golden Cheetah (app)	—	●	●	—
	Google Fit	●	○	○	—
	GPS Foot Pedals	●	—	—	—
	Pedometer	●	—	—	—
	Power Meter	●	—	—	—
	Strava (app)	●	●	●	●
	Suunto Watch	●	○	●	○
	Whoop Band	●	●	●	○
Maintainers	Apple Health (app)	○	●	○	●
	Apple Watch	●	●	●	●
	Fitbit Watch	●	●	○	●
	Garmin Connect	○	●	○	●
	Garmin Vivoactive Watch	○	●	○	●
	Kronos Watch	●	●	●	●
	Lose It (app)	○	●	○	—
	My Fitness Pal (app)	○	●	○	○
	My Maintainers Pal (app)	○	●	○	○
	Veryfit Watch	○	●	○	—
	Virgin Pulse (app)	○	●	○	—

Note: "●" means feature was being used by at least one participant, "○" means features was not being used by participants who reported using the wearables/applications, "—" means device does not have such a feature.

to other analysis tools. In contrast, ***maintainers*** P15 pointed out that the Apple Watch's three rings, representing exercise progress, were crucial for tracking their daily achievements.

The analysis also found that participants actively used various personal informatics tools, selecting specific tools for different data collection and goal-tracking needs. For instance, P2 mentioned using the smartphone application *Lose It* for weight management, *My Fitness Pal* for nutrition, and *Apple Watch* for activity tracking, showcasing the diversity of tools even when there was feature overlap. P3, for instance, solely used his smartwatch for data collection, although it had visualization features. He preferred to export the collected data to another online analysis tool for data analysis and visualization, as he believed it provided more valuable insights into his running activities.

The interviews also delved into the lifecycle of personal informatics tools. As expected, tools were most commonly discontinued when users acquired new, more capable tools, such as upgrading to the latest fitness watch. Many users considered the value of a tool's data in connection with its ease of use when deciding to discontinue its use. For example, P4 noted, *"I just find [the chest strap heart rate monitor] to be a little bit too cumbersome, so I stick to an accurate enough sensor on my watch."* Usability issues and insufficient battery performance were also common reasons for discontinuation, aligning our findings with Lazar et al.'s report on adoption and abandonment behaviors. Resulting in missing data in their records.

All participants in our study reported that personal informatics tools had a positive impact on their goals. Many, like P12, felt that the tools made them more accountable, noting, *"I'd be lost without [the Fitbit], I wouldn't have any idea of [goal progress]."* Participants P6 and P20 valued alerts about not meeting daily and weekly targets, as these alerts triggered self-reflection on their decisions and behaviors. Similarly, P5, P9, and P12 all expressed satisfaction in seeing their workout results quantified, emphasizing the value of knowing their numerical caloric achievements while performing various activities. Indicating the importance of having complete data in achieving their goals.

Table 10: Reflective behaviors discovered in the semi-structured interview study.

Category	**Motive of use**	**Data Usage (goal)**	**Frequency of use**		**Features used**
			Collection	Reflection	
Trainees	Cycling Marathon Martial arts Coach others Weight lifting Cycling tournament Running tournament	Training power Training strength Training sprint Analyze performance Balance training load Inform training plan Optimize performance Improve performance Maximize abilities in running Replicate best performance	Upon training	Daily, upon training, weekly, monthly	Identify important moments; Training stress balance; Chronic training load; Elevation gain; Speed; Power; Heart rate zones; Cadence; Pace; Training intervals; Identify min/max value for each feature; Recovery score.
Maintainers	Educate self Log activities Improve health Maintain health Manage weight Maintain heart health	Map routes Plan meal Monitor Sleep Lose weight Maintain weight Check Mileage Regulate heart rate Check step counts Check Calories burned Monitor Sleep conditions Log Swimming hours	Daily	Weekly, monthly	Calories intake and nutrition; protein and carbohydrates in each meal; Active hours; Sleep duration per day; Total workout hours; Check completion rate; Check targets set for activities; Compare performance.

4.2.2 Users' expectations of consistency and completeness in PI data

Exploring how individuals employ their PI tools for various purposes (as discussed in Section 4.2.1, table 10 offers additional insight into participants' usage patterns by distinguishing them as ***trainees*** and ***maintainers***. We examined the features and utilization frequency in two stages: 1) data collection, involving the acquisition of sensory data through PI tools, and 2) data reflection, which entails analyzing the gathered data to extract actionable insights for behavior change.

The analysis of usage frequency and the specific features employed during reflection is derived from participants' explicit descriptions of their reflection processes. The outcome of this analysis is presented in the last two columns of the table. A noticeable divergence was observed between the distinct user classes. ***Maintainers*** displayed a consistent pattern of utilizing and actively engaging with the tools at the time of data collection, frequently to confirm data acquisition and monitor progress toward their goals. On the other hand, ***trainees*** exhibited lower engagement, often checking the tools just once a day or periodically throughout the week to reflect on their activity trends and broader health maintenance goals.

The reflection behaviors among ***trainees*** predominantly revolved around the concepts of *understanding the past* and *predicting the future*. For instance, P1, who falls into this category, emphasized the importance of utilizing past performance data to guide adjustments in a training routine, stating, *"if somebody is losing races and their sprint is not as good as before, he/she will need more sprint work (recommended method of choice for cardiovascular exercise)."* Similarly, P3, another ***trainees***, explained how this reflective process helps predict race performance, noting, *"it helps you predict how fast you can actually race because you know exactly how long you can hold at a certain heart rate for."* Additionally, P13, yet another ***trainees***, expressed the need to maintain a reasonable balance of training stress, stating, *"based on how recovered I am, this is how much strain I should put on my body."*

On the other hand, ***maintainers*** typically engaged in reflection to assess their past behaviors in relation to their goals while staying focused on the present. For instance, P6, a maintainer, explained their reflective process by saying, *"usually just checking how far I am from my target."* Additionally, 7 out of the 10 maintainers, namely P5, P6, P9, P4, P13, P8, and P20, mentioned that they primarily logged details about the number of exercises or the frequency of exercise sessions.

They did this to track their progress toward achieving the *perfect week*, which involves meeting specific daily, weekly, and exercise-related goals.

Differences in the value of precision and detail were also evident. ***Trainees*** were inclined to seek specific insights and compare precise data across various timeframes (e.g., daily, weekly, and monthly). For instance, trainees found value in moments of significance, chronic training loads, training intervals, and recovery scores, using them to strike a balance between light and intense training. P3 mentioned that having access to the minimum and maximum values of these features allows them to *"analyze from an analytical perspective about what my body is capable of"*. He even suggested that replicating a certain heart rate and pace could help mimic past successes in a new competition.

Conversely, ***maintainers*** focused on obtaining a summary of insights that encompass a wide range of past activities. The last column in the table shows that features such as active hours, sleep duration, total workout hours, and more were oriented toward tracking general trends and broader health goals.

4.2.3 Missing PI data conflict and users' expectations

The reflective behavior discussed in section 4.2.2 relies on the PI data that is collected. We observed that missing data in either of the stages can lead to problems. For ***trainees***, the absence of data for certain activities hinders effective comparisons across activities or tracking specific performance metrics. For ***maintainers***, the lack of data might incorrectly suggest that short-term or long-term goals are not being met. In both groups, the absence of essential data was associated with the abandonment of devices, indicating a low tolerance for accommodating the functional limitations of the tools.

Both ***maintainers*** and ***trainees*** reported encountering missing data while using the PI tools in our study. We conducted an analysis to understand the factors that contribute to missing data and its impact throughout the personal informatics lifecycle. Our analysis identified two dimensions: 1) human-related factors, such as missing data resulting from participant errors or behaviors, and 2) device-related factors, such as missing data due to malfunctions or misconfigurations of wearables or applications. These dimensions are applicable to both groups.

Human-related causes of missing data include:

- *Failure to initiate data collection*: For instance, participants may forget to activate the application or device to collect data for a particular activity. This may also involve forgetting to annotate data or manually enter information.
- *Failure to carry the device*: Participants might neglect to bring or wear the device for an activity. In some cases, like that of P4, a ***maintainer***, and P8, another ***maintainer***, who forgot to take their smartwatches on vacation, these gaps can extend over several days. Most commonly, participants tend to forget to put the device back on after charging.

Causes of missing data related to devices include:

- *Battery depletion*: A significant number of participants encountered battery-related issues during their activities. For example, P8, a ***maintainer***, mentioned, *"sometimes I forget to charge my watch, it'll die in the middle of a workout or run."*
- *Device malfunction or limitations*: P6, a ***maintainer***, reported that her smartwatch failed to count swim laps accurately.
- *Syncing issues*: P7, a ***trainee***, noted that when syncing data to other applications, data points were lost due to synchronization problems.
- *Device precision*: All participants indicated that their devices often failed to capture activities with the correct level of precision. For instance, P4, a ***maintainer***, mentioned that sometimes, when holding the wheel while driving, the smartwatch would register the vibration as steps. P16, a ***trainee***, stated, *"sometimes, if I'm walking around the class and talking loudly, my heart rate might spike a little, but the smartwatch might count it as an active workout, which it is not."*

Quantitative analysis of the 5-point Likert scale ratings (questions provided in Table 11) revealed that 40% of maintainers and 40% of trainees indicated that missing data had no impact on their goals (Q4, rating "Not at all"). Additionally, 60% of maintainers and 40% of trainees reported that missing data slightly affected their goals (Q4, rating "Slightly"), 20% of maintainers stated that missing data somewhat influenced their goals (Q4, rating "Somewhat"), and 20% of maintainers expressed that missing data had a moderate to extreme impact on their goals (Q4,

Table 11: Quantifiable questions used in the semi-structured interview study.

Questions	Ratings
Q1. Please rate your frustration level when the data is missing.	“Not at all”
Q2. Please rate your frustration level when the data is inaccurate.	“Slightly”
Q3. Please rate your trust level towards your tracking data.	“Somewhat”
Q4. Please rate the level of influence missing data has on your goal.	“Moderately”
	“Extremely”

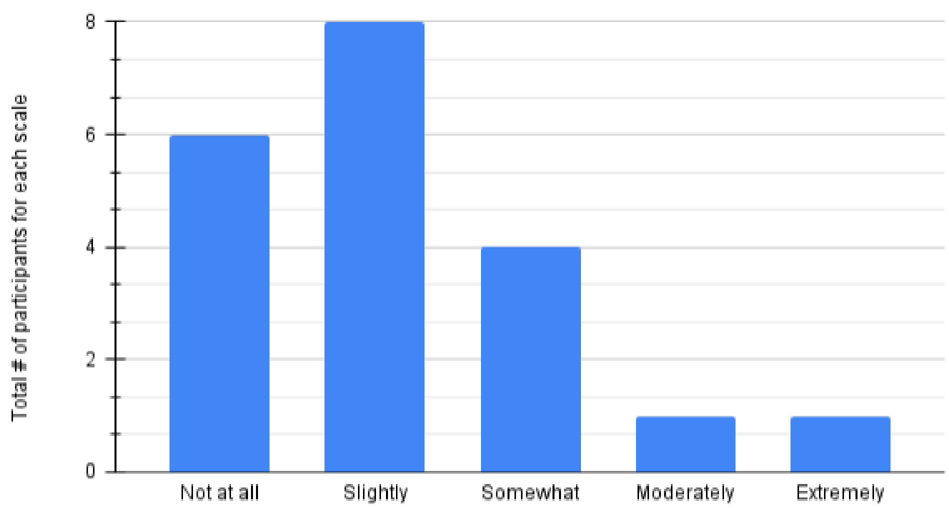
rating “Moderately”). Figure 8 visually depicts the extent to which missing data affects their goals as per the Likert scale ratings.

20% of maintainers and 30% of trainees mentioned taking action when they noticed missing data. A similar percentage of participants stated that they would refer to their friends’ data if it were available for estimating their own PI data when data was missing. Additionally, 20% of trainees mentioned that they would estimate the data based on how their body feels. During the reflection stage, when participants noticed that the tools did not capture their hours or activities during the day, they would sometimes redo the workout to make up for the missed data. For estimated data, they would reflect on it in a similar way as regular data. However, some participants also preferred having no data to having inaccurate data under certain circumstances. For example, P18, a trainee, mentioned, *“The estimated calories burned are accurate to about 90%, which is fine. But for the estimated heart rate? No, absolutely not. I would rather the app (Google Fit) tell me it did not capture it (heart rate).”*

Interestingly, some participants were motivated to repeat activities in order to capture the PI data required to achieve their goals accurately. For example, P8, a maintainer who stated that missing data somewhat affects her goal, mentioned, *“If I set 100 minutes, and I had missing data, and the watch shows I would have gotten 30 minutes. I’ll probably just run again with my watch to get more minutes so I can reach my goal set on the app.”*

The participants’ levels of frustration with missing data varied. Specifically, 10% of trainees and 20% of maintainers reported feeling somewhat frustrated when their data was missing (Q1, rating of “Somewhat”). Additionally, 30% of trainees and 20% of maintainers indicated that their frustration towards missing data was moderate (Q1, rating of “Moderately”), while 30% of trainees

Figure 8: Effect of missing data on goal.



Note: The figure presented illustrates the impact of missing data on participants' training goals from the interview study.

rated their frustration as extremely (Q1, rating of "Extremely"). An equal percentage of 20% in both groups expressed a slight degree of frustration (Q1, rating of "Slightly"), and 10% of trainees and 30% of maintainers reported that they were not at all frustrated about missing data (Q1, rating of "Not at all"). Figure 9 visually represents the extent to which missing data affects their frustration level based on the Likert scale ratings.

All participants acknowledged that the data collected had some discrepancies, and they recognized that it is not technically practical to have data with 100% accuracy. For example, P5, a maintainer, commented, "I like things to be accurate. I won't say I'm a numbers person per se, but I like things organized and proficient, so inaccurate data did bother me at first." While acknowledging that PI data is not 100% accurate, overall levels of trust in the collected PI data were high. Specifically, 50% of trainees and 30% of maintainers reported an extremely high level of trust in their data (Q3, rating of "Extremely"). Meanwhile, 30% of trainees and 50% of maintainers expressed a moderate level of trust (Q3, rating of "Moderately"), and 20% of trainees and maintainers in each group reported a somewhat level of trust in the data (Q3, rating of "Somewhat"). Figure 9 visually illustrates the extent to which missing data affects their trust level based on the Likert scale ratings.

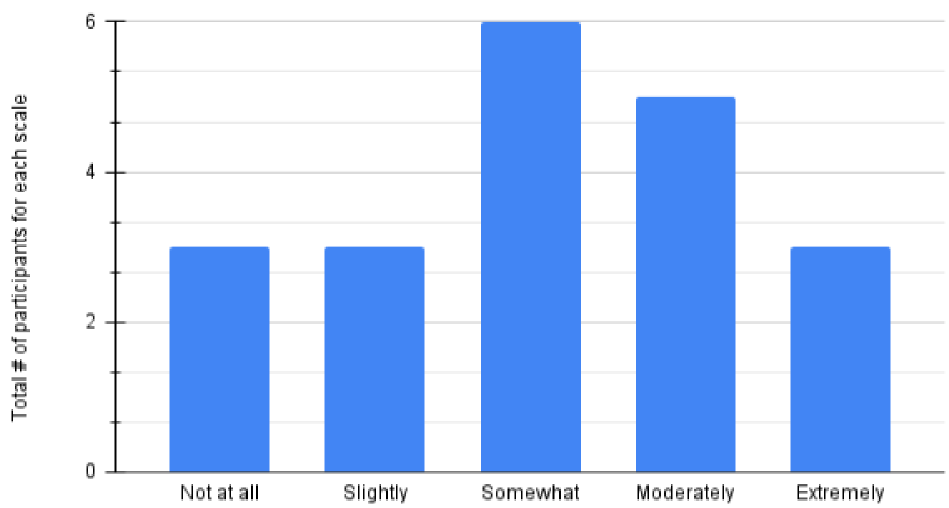
Upon analyzing the data, we observed that missing data has a negative impact on participants in both groups. However, there are slight differences in the level of effects. For instance, participants in the trainees' group tend to experience more frustration when their data is inaccurate. In contrast, the maintainers' group shows a higher degree of frustration when the data is missing.

4.3 Observations

Through the semi-structured interview, we found different usage patterns for trainees and maintainers; we identified that missing data causes conflicts. We also elaborated on how their expectations were different between these two groups.

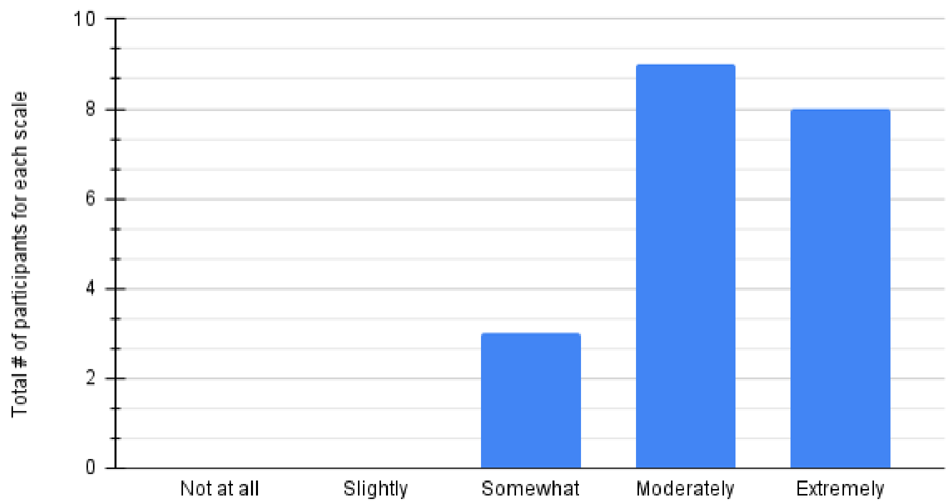
Our findings highlighted the significance of forgetting to wear or charge devices and how it affects the value and utility of personal informatics (PI) tools. This presents a *fundamental* design challenge for PI technology, as there is no foreseeable technological advancement where these

Figure 9: Frustration level for missing data.



Note: The figure presented illustrates the level of frustration experienced by participants when their data is missing from the interview study.

Figure 10: Trust level for wearable data.



Note: The figure presented illustrates the level of trust that participants have in their data from the interview study.

tools and devices automatically charge and attach to users. Addressing this challenge is crucial for ensuring the long-term use and utility of PI. Therefore, in this section, we explore the question of how the design of PI tools can be expanded to assist users in reflecting on and addressing missing data in their tracking activities.

In this context, we propose adopting the concept of synthetic data as a guiding principle in the design of personal informatics (PI) tools. Synthetic data refers to the creation of visual representations to aid users in distinguishing between missing data and cases where no data was collected. It also involves helping users estimate their missing data as part of the reflection process.

Our results pointed to several key implications for enhancing PI tools by incorporating synthetic data into the user experience. Integrating synthetic data (missing data estimation) as a design principle for PI tools means that the tool should not only address where and how to incorporate it based on various user motivations but also focus on how to represent it within the tool visually.

4.3.1 Usage Behavior

Our study identified three primary usage behaviors: 1) *understanding the past*, 2) *knowing the present*, and 3) *predicting the future* (see Section 4.2.2). These behaviors are crucial for self-reflection and obtaining actionable insights for behavior change. However, our analysis revealed that existing tools do not adequately support the integration of these three usage behaviors. Consequently, participants often use multiple PI tools for different aspects of their tracking needs. To address this issue, tools designed to incorporate synthetic data should consistently address three key aspects: 1) where synthetic data is present in weekly, monthly, and yearly data; 2) how synthetic data is utilized for goal-tracking estimation; and 3) how users can control the use of synthetic data to uncover patterns that inform future behaviors and goals. These three usage behaviors are interrelated, and providing a consistent representation of how missing data is handled and interpreted can reduce fallacious insights and enhance users' confidence in their data.

4.3.2 Sensitivity to Missing Data

The results indicated varying levels of tolerance to missing data among participants (refer to Section 4.2.3). For instance, some participants mentioned they could tolerate one or two days of missing data per week, while others had a much lower tolerance and could only accept a few hours of missing data each day. Participants consistently demonstrated tolerance for missing data in the context of how it interfered with their tracking goals, trends, and specific activities. This finding highlights that there is no one-size-fits-all solution to address the sensitivity to missing data using a single mitigation approach.

Based on feedback from participants, three design approaches are suggested:

- **Manual Data Entry**: Allow users to input missing data manually; this is ideal for small data gaps that correspond to specific activities, such as an outdoor run.
- **Algorithmic Estimation**: Implement algorithmic methods to estimate missing data gaps based on prior data collection; this is well-suited for larger data gaps in daily, repetitive activities, like determining the number of steps taken during a routine evening walk.
- **Event Based Estimation - Algorithmic Estimation with User Guidance**: Combine algorithmic estimation with user interfaces that assist users in refining estimates based on known conditions of the missing data; this approach is useful for more extensive data gaps that involve specific activities or sets of activities, such as estimating the performance of a long-distance run by considering start and stop times and locations.

Further research is necessary to assess the effectiveness of each approach and determine the conditions under which users prefer one method over the others.

4.3.3 Visual Representation of Missing Data

In our study, all the personal informatics (PI) tools used by participants (as listed in Table 9 were found to handle missing data by treating it as the absence of data. In cases where an exercise or activity was performed, but data on the event was not collected, these tools typically communicate to the user that no activity took place. This can be misleading and has the potential to impact a user's personal health and fitness goals inaccurately.

To address this issue, PI tools should offer a more distinct visual representation to help users differentiate between instances where data is missing and moments when activity levels are genuinely low. Recognizing gaps in data collection when a device was worn can assist users in placing the data in context, enabling them to better reflect on their physical activities. Moreover, visualizing these gaps can provide a natural opportunity in the user experience for users to manually estimate missing data or allow an algorithm to generate synthetically derived data based on heuristics.

4.3.4 Trust in Data

Our analysis revealed that participants generally have a high level of trust in the accuracy of the personal informatics (PI) data they collect, as discussed in Section 4.2.3. This trust is crucial for maintaining adherence to and continued use of PI tools. Participants often sought to validate this trust by comparing their data with other sources, whether through direct comparisons with individuals who performed the same activity, real-time examination of maps during runs or walks on specific routes, or estimating data based on their own perception of physical fatigue.

This behavior suggests that tools designed to enable the creation of synthetic data should allow users to:

- View synthetic data within the context of plausible and previously observed behaviors to ensure that synthetic data is grounded in real data (e.g., comparing to data from the previous five activities).
- Group and label synthetic data as specific, personally relevant events (e.g., labeling an activity as 'run with John').
- Have the ability to manipulate and adjust algorithmic estimates based on their intuition or preferences for how missing data should be generated.

4.4 Discussion

In this study, we explored a novel perspective, namely the concept of missing data, within the realm of personal informatics. We examined how the absence of data affects user behaviors and perceptions in real-world, day-to-day usage scenarios. Our research expanded the existing knowledge within the community by shedding light on the various situations that lead to missing data and the subsequent impact on data collection and reflective behaviors.

Our analysis revealed the challenges arising from missing data, encompassing both individual and technical issues. Furthermore, it offered valuable insights into how personal informatics tools should evolve to accommodate synthetic data and provided guidance on the strategies that can be employed to facilitate this transition.

Through the analysis, we identified two distinct groups: trainees and maintainers. Trainees predominantly employ their personal informatics tools to *understand their past* and *predict the future*. On the other hand, maintainers primarily utilize their PI tools to *know the present*.

For both of these user groups, we have presented a comparative examination of how missing data influences the life cycle of personal informatics and lessons to be learned from the perspectives of both trainees and maintainers. Our analysis has unveiled significant limitations in existing tools, particularly the absence of adequate representation of missing data. In response, we have outlined design principles to serve as guidelines for future tools, enhancing the overall user experience with personal informatics data.

For future research, it will be essential to employ diverse methodologies for generating and representing synthetic data in personal informatics tools. Conducting user studies is crucial to assess the effectiveness of each approach and to determine the specific conditions under which each approach is preferred by users. This empirical investigation will provide valuable insights into the practical implementation of synthetic data in PI tools and help refine the design principles proposed in this study.

5.0 System Implementation

In this chapter, we introduce MissFit, a web application specifically developed to address the issue of missing data within Personal Informatics (PI) systems. MissFit leverages data from Fitbit devices and engages with users to fill in the gaps. It provides estimates for missing daily steps through six distinct designs, which include two scenarios and three methods. The scenarios, derived from common causes of data omission (see Section 4.2.3), are: Scenario A, which accounts for missing an entire day's data, and Scenario B, which pertains to missing data from one or multiple activities within a day. The three methods, inspired by typical reactions to missing data (see Section 4.3.2), are: manual input, algorithmic estimation, and event-based estimation. The manual input method enables users to directly enter missing data, the algorithmic method estimates gaps using historical data, and the event-based method allows users to select similar events—considering aspects like neighborhood, frequency, duration, mileage, and routes—to estimate missing data.

It's worth mentioning that we opted for these metrics because we found them to be most relevant for individuals who use their tracking devices primarily for health maintenance (see Chapter 4). While other metrics, like heart rate, is available in some fitness trackers/devices, we decided against including them in MissFit. This decision was based on our semi-structured user study findings, where maintainers (see Section 4.2) expressed limited interest in heart rate data. Since MissFit is designed with maintainers in mind—owing to their larger user base—we aimed to cater to a wider audience. Our choice of metrics is informed by our research to date, but this does not preclude the possibility of exploring other metrics in future interfaces. Indeed, MissFit's approach to eliciting missing data could apply to a variety of metrics.

The MissFit web app is composed of two main components:

- *Graphical User Interface (GUI):* The user interface provides the interaction mechanism for users to leverage the underlying functionalities. Specifically, the user interface allows users to select scenarios and estimation methods they wish to go through to recover the missing data or activities. It will also display an interactive bar chart with their exercise data (including the missing data) for the past two weeks. After finishing the steps for the selected estimation method, the estimated data will be displayed to users.

- *Data Server:* The data server was designed to fetch, process, and store the Fitbit exercise data so that MissFit can access data seamlessly. This process involves generating an access token to retrieve Fitbit data, sending requests to the Fitbit company to obtain the data, storing the data in a JSON file, and placing the JSON file on the JSON server so that MissFit can fetch data from this server.

In MissFit, we use bar charts to represent users' data, with design references to current industry products like Fitbit and iWatch. This approach aligns with the design implication proposed by Cuttone et al. [30] that PI systems should make data interpretable at a glance (see Section 2.4 for more details).

5.1 Graphical User Interface (GUI)

The user interface visually represents the daily step counts for the past two weeks, highlighting instances of missing data or activities. It showcases two scenarios and three methods available to the user, indicating the steps they must complete to obtain their estimated data. There are six methods: Scenario A and Scenario B, each with three associated methods.

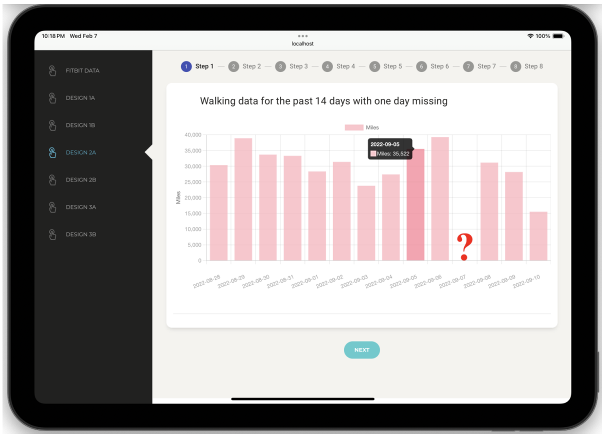
As shown in Figure 11a, the interface comprises two main components: Estimation Method Control (the left panel) and Estimation Methods (the right panel).

5.1.1 Estimation Method Control

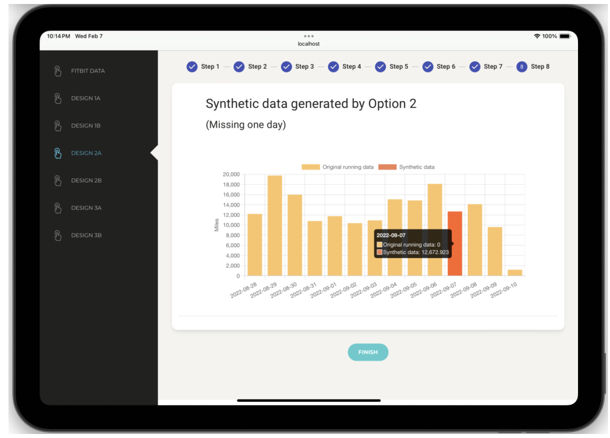
This control enables users to choose a scenario and method that best suits their circumstances for estimating missing data (see Figure 12). The control is automatically displayed on the left side of the app when it is running and at every step. This placement was selected to reinforce the concept that it is a web application, enhancing intuitiveness for users as they navigate and interact with the interface. Additionally, it provides quick access to the available options that users can utilize to estimate their missing data.

The estimation method control consists of six options (Figure 12):

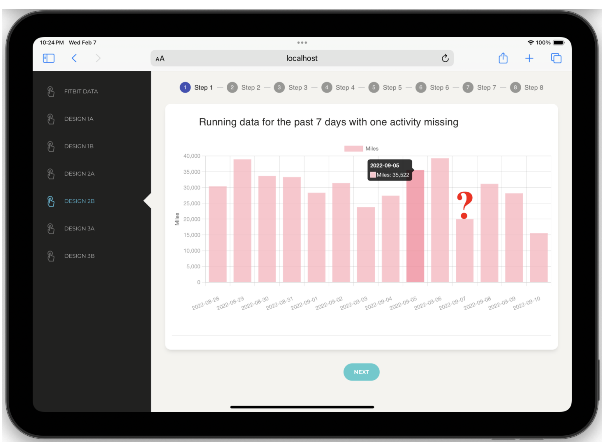
Figure 11: Commonly used user interface in MissFit.



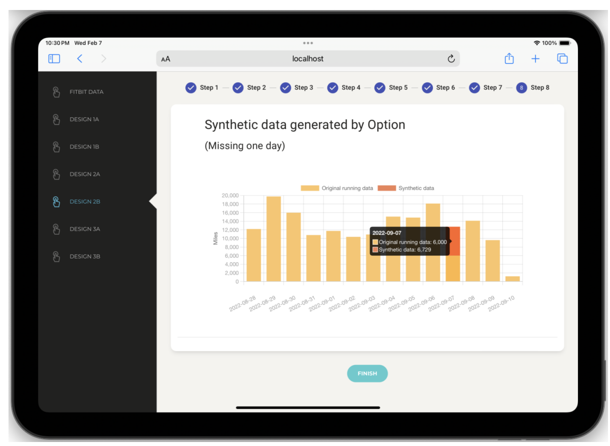
(a) Daily step data with one day missing.



(b) Daily step data with estimated data.



(c) Daily step data with activities missing



(d) Daily step data with estimated data

Note: The screenshot illustrates the user interface commonly used in all methods. Screens (a) and (c) represent users' initial screens when selecting a specific method on the left. In (a), the most recent daily step data that is missing for an entire day is displayed, while (c) shows missing activities. Screens (b) and (d) represent the final screens a user will see when completing all the required steps for the method. These screens display the estimated data using the selected method, with the red bar indicating the estimated data. Screen (b) corresponds to missing an entire day, while (d) corresponds to missing activities.

- *DESIGN 1A*. If the user misses one day of step data and wants to get the estimated data with minimal effort. For example, if the user forgot to put their Fitbit device back on after charging but still engaged in regular exercise activities, the entire day's activities were not recorded. They may want to estimate the missing data with minimal effort after a tiring workday. In such cases, the user can choose to use this method.
- *DESIGN 1B*. If the user is missing some portion of the daily step data and has confidence in algorithms in general, for example, if the Fitbit device runs out of battery in the middle of the day, leading to some activities not being captured, and they wish to obtain estimated data quickly, the user can choose this method.
- *DESIGN 2A*. If the user is missing an entire day of step data and can recall daily events over the week—for example, if the user forgot to sync their Fitbit device, resulting in a missing day in the record, but has a good memory of their daily activities—then the user can use this method to retrieve their missing data.
- *DESIGN 2B*. If the user is missing some portion of the daily step data and enjoys comparing trends/similarities in their weekly data, for example, if the Fitbit device runs out of battery in the middle of the day and they notice that the missing data is similar to Monday and Friday's data, the user can choose this method.
- *DESIGN 3A*. If the user is missing an entire day of step data and is aware of the steps they missed, for instance, if the user engaged in the same activities as their friends during the missing day and can obtain their steps from their friends, they can use this method.
- *DESIGN 3B*. If the user is missing one activity of step data and knows the steps from that activity, for example, if the user typically runs or walks around their neighborhood following the same route and forgot to bring their Fitbit device one time but knows the steps because they are familiar with the route, then they can use this method.

Offering all method options is necessary, as we found that users have different reasons that cause missing data in their wearable devices and utilize various methods to estimate the missing data themselves (see Chapter 4, Section 4.2). For instance, a user who bikes with a group of people but experiences a drained battery in the middle of the trip may refer to his friends' data to retrieve information about his cycling trip.

Figure 12: Estimation methods and designs in MissFit.

Note: The figure presented illustrates the available estimation methods and designs. The Estimation Method Control is displayed on the left panel for each method and step, allowing users to configure the scenario and method for estimating their missing data. 'DESIGN 1A' represents Scenario A and utilizes an algorithmic approach for estimating the missing data. 'DESIGN 1B' represents Scenario B with an algorithmic approach. Using an event-based approach, 'DESIGN 2A' and 'DESIGN 2B' correspond to Scenarios A and B. 'DESIGN 3A' and 'DESIGN 3B' under Scenarios A and B utilize a manual-input approach for estimating the missing data.

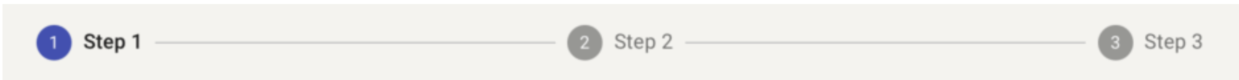
5.1.2 Estimation Methods

The estimation method provides steps that users must follow when estimating their missing data. Upon selecting a method, an interactive bar chart representing the most recent fourteen days of daily step data is displayed, with a question mark indicating the missing data (see Figure 11a) and partial daily data (see Figure 11c). Each estimation method has specific steps the user must follow before obtaining the estimated data; the stepper bar is located on the top right of each page (i.e., see Figure 11a for the location on each page). Figure 13 shows all examples of how the stepper bar indicates different statuses when the user uses a specific method. After completing the steps for each method, the user is presented with another interactive bar chart containing the estimated data (see Figure 11b and 11d).

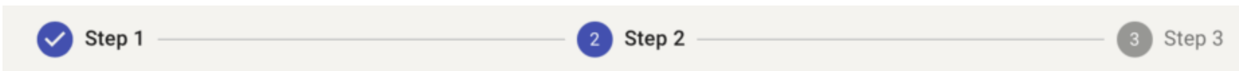
There are three methods made available to users:

- **Algorithmic Method.** *DESIGN 1A* and *DESIGN 1B* (see Section 5.1.1) utilize this method. When the user uses the algorithmic method to estimate their missing data, the method will briefly explain through a short text, followed by the generation of estimated data. Figure 14 showed the process of using 'DESIGN 1A' to estimate their missing data. This method is automated and requires minimal effort from participants. It is designed for cases where users trust algorithms and seek a fast and efficient solution.
- **Event-based Method.** *DESIGN 2A* and *DESIGN 2B* (see Section 5.1.1) utilize this method. When the user uses the event-based method, there are eight steps the user needs to complete. As shown in Figure 15, the user will see the screenshot in (Step 1) when they select the 'DESIGN 2A'. Clicking the 'NEXT' button will lead them to (Step 2) which asks them to choose the neighborhood where they walked on the missing day. For example, if they cycled in Bloomfield on the missing day, they would need to select Bloomfield. (Steps 3, 4, and 5) require users to select similar days based on duration, mileage, and frequency. In (Step 6), users are presented with the four most recent routes they walked and asked to choose the route closest to the missing day. (Step 7) provides a summary of their selections in previous steps, and (Step 8) displays the user's estimated data calculated from their input. This method is designed for cases where users have repetitive exercise routines and enjoy comparing trends or similarities in their data.

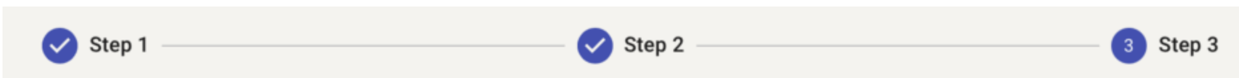
Figure 13: Stepper status in MissFit.



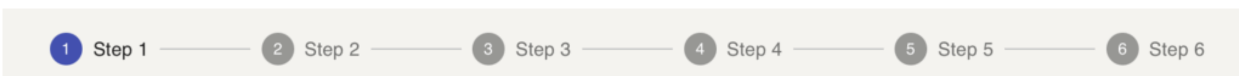
(a) Stepper status when user clicks on DESIGN 1A, 1B, 3A, and 3B.



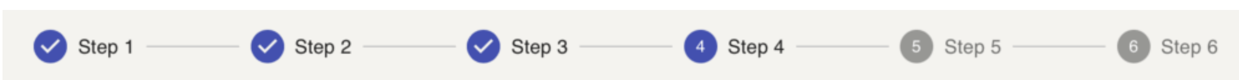
(b) Stepper status when the user half of the steps for DESIGN 1A, 1B, 3A, and 3B.



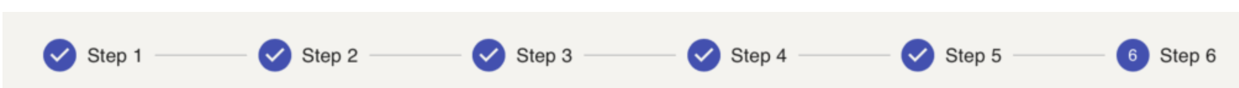
(c) Stepper status when the user finishes all the steps for DESIGN 1A, 1B, 3A, and 3B.



(d) Stepper status when user clicks on DESIGN 2A and 2B.



(e) Stepper status when the user finishes half of the steps for DESIGN 2A and 2B.



(f) Stepper status when the user finishes all the steps for DESIGN 2A and 2B.

Note: The screenshot illustrates the stepper status under different circumstances. The stepper conveys three pieces of information to users: 1) which step they are at, 2) how many steps they have finished, and 3) how many steps they have in total. For example, in (a), there are a total of three steps, and the user is currently on 'Step 1'. In (b), the user has completed 'Step 1' and is now on 'Step 2'. And in (c), the user is on 'Step 3', having finished all prior steps, and 'Step 3' is the final one.

- **Manual-input Method.** *DESIGN 3A* and *DESIGN 3B* (see Section 5.1.1) utilize this method. There's only one step the user needs to complete. As shown in Figure 16, if the user is missing one day of their step data, they can directly input the information into the textbox (see Figure **??**). If the user is missing one activity, they can provide detailed information about the activity in the textbox (see Figure **??**). The method will generate the estimated data based on the input. This method is designed for cases where users have good estimates themselves about the missing data or in situations where they can refer to their friends' data.

5.2 Data Server

MissFit utilizes the data server as an intermediary between MissFit and Fitbit exercise data for fault prevention. To prevent abuse, Fitbit has imposed rate limits on their API calls. For example, if a user mistakenly sends repeated fetch data requests that exceed the maximum number of requests an application can make within a specific time period, Fitbit will decline any request that comes from that application for a period of time. Since MissFit uses different data sources, it needs to send multiple requests to Fitbit, making it more sensitive to Fitbit's rate limits policy. Using a data server helps mediate this issue.

As shown in Figure 17, the data server comprises three main components: Access Token, Refresh Token, and JSON Data Storage.

5.2.1 Access Token

An access token is a credential that authenticates and authorizes access to protected resources in the API or web application. In the Fitbit Web API context, an access token is generated when a user (any user with a Fitbit account) grants permission to a third-party application (MissFit web app) to access their Fitbit data. This access token is then used in subsequent API requests to validate the identity and permissions of the requesting application.

Obtaining an access token involves an OAuth 2.0 authentication process, where the user is redirected to the Fitbit authorization server, prompted to log in, and granted permission to the

Figure 14: Steps for algorithmic method in MissFit.

Note: The figure shows the steps the user goes through when using the algorithmic method (i.e., DESIGN 1A) to estimate their missing data. The user will see (Step 1) when selecting ’DESIGN 1A’. After clicking the ’NEXT’ button, the user will move to (Step 2). Clicking ’NEXT’ again leads to (Step 3). And clicking the ’Finish’ button brings the user back to (Step 1).

Figure 15: Steps for event-based method in MissFit.

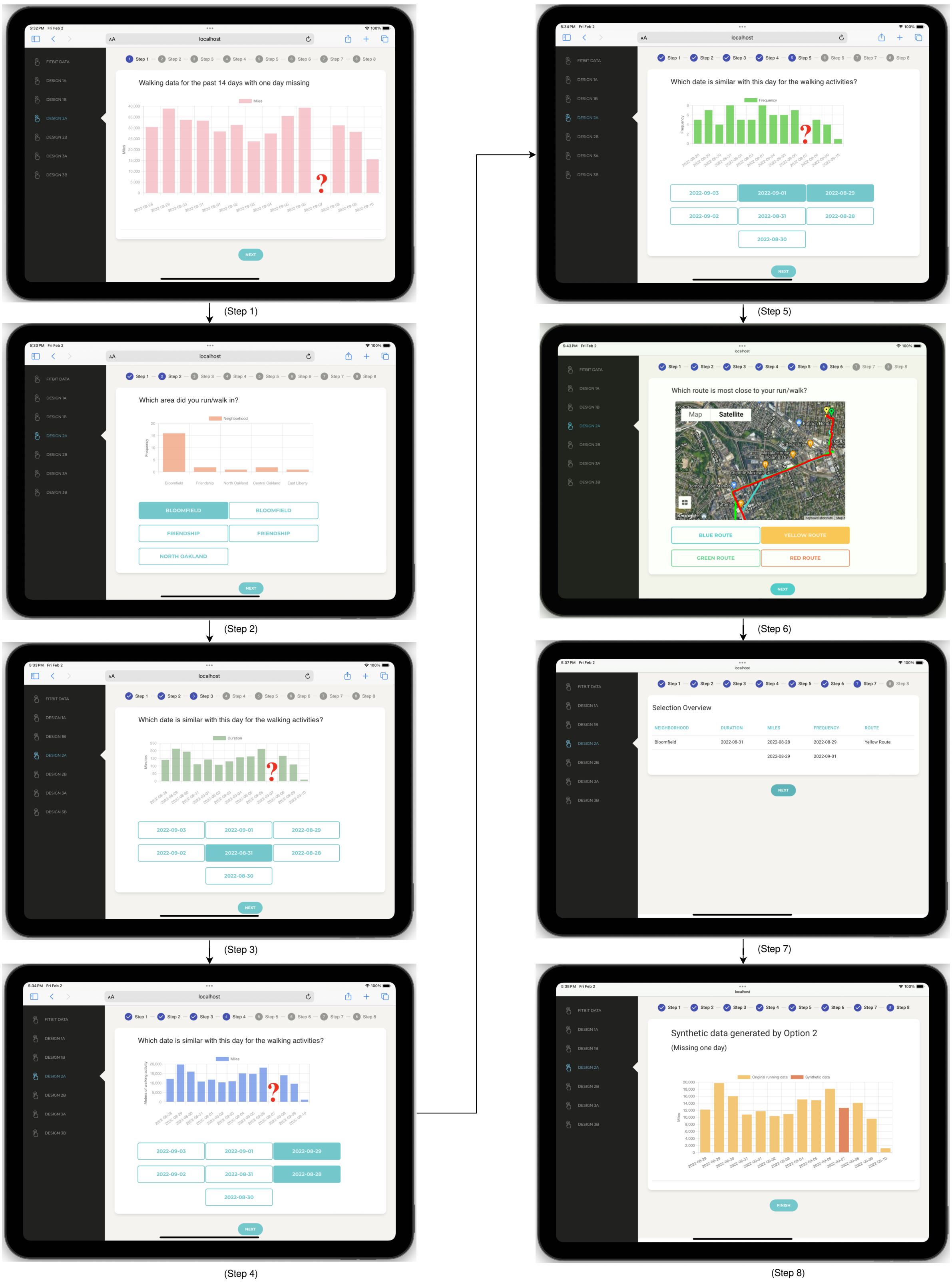
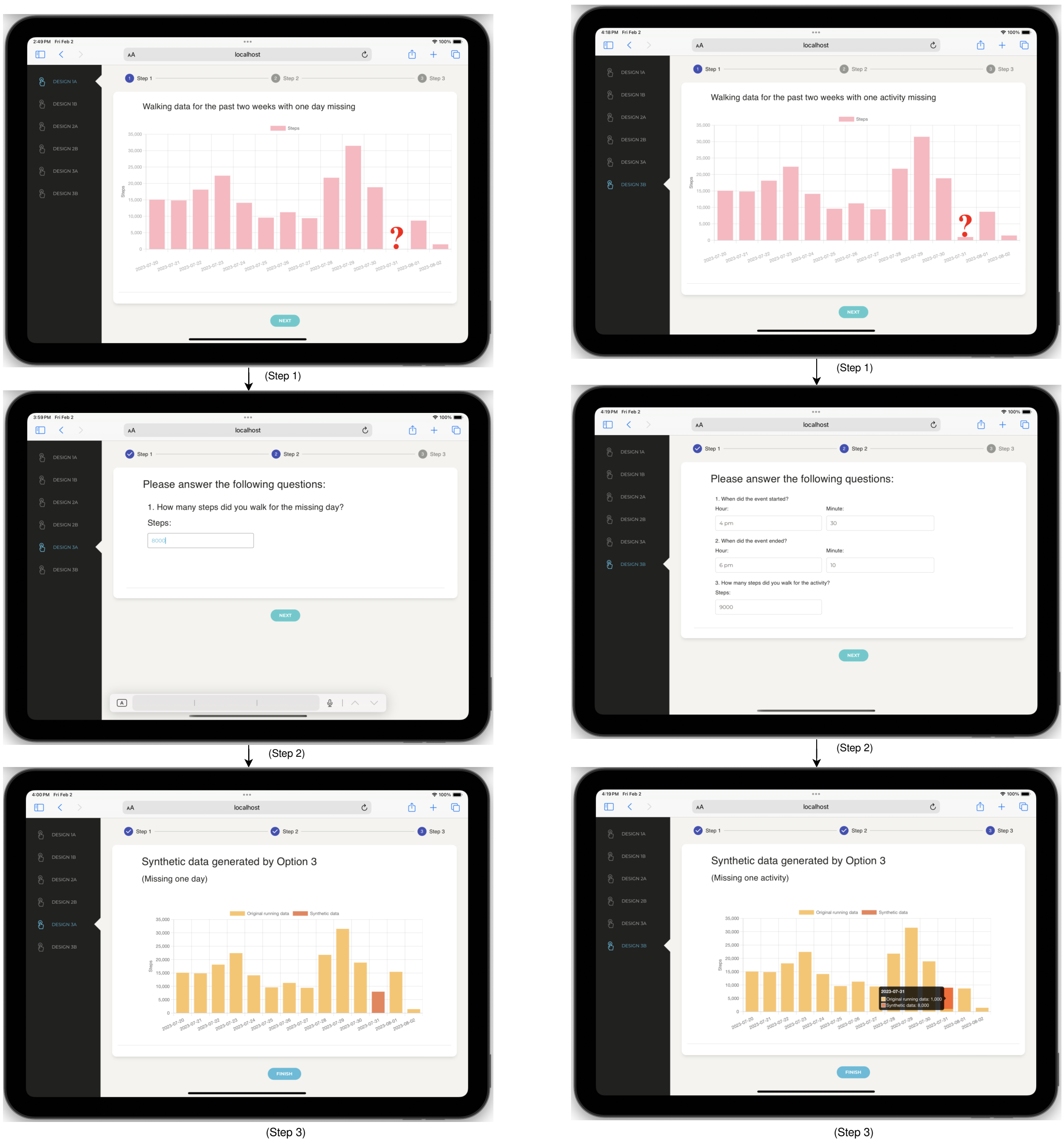
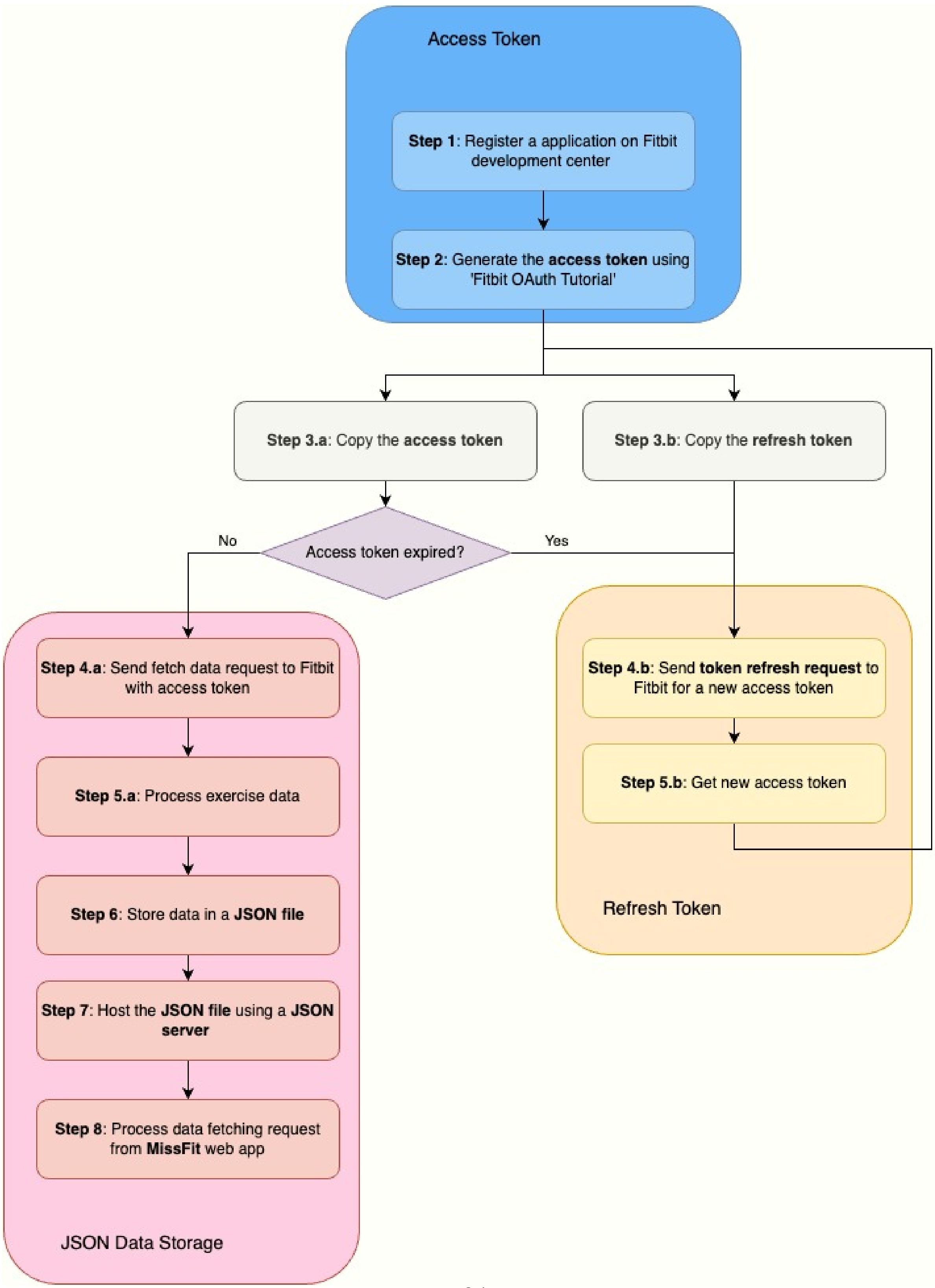


Figure 16: Steps for manual input method in MissFit.



Note: The figure illustrates the user's steps when using the manual input method to estimate their missing data. In 'DESIGN 3A' (left), the user will encounter (Step 1) upon selecting 'DESIGN 3A'. After clicking the 'NEXT' button, the user will move to (Step 2). Clicking 'NEXT' again leads to (Step 3). Clicking the 'Finish' button brings the user back to (Step 1). 'DESIGN 3A' (left) has the same process as 'DESIGN 3B' (right).

Figure 17: Flow chart and components for the data server in MissFit.



application. Upon approval, the Fitbit server issues an access token to the application, which is included in subsequent API requests as an authorization header or parameter.

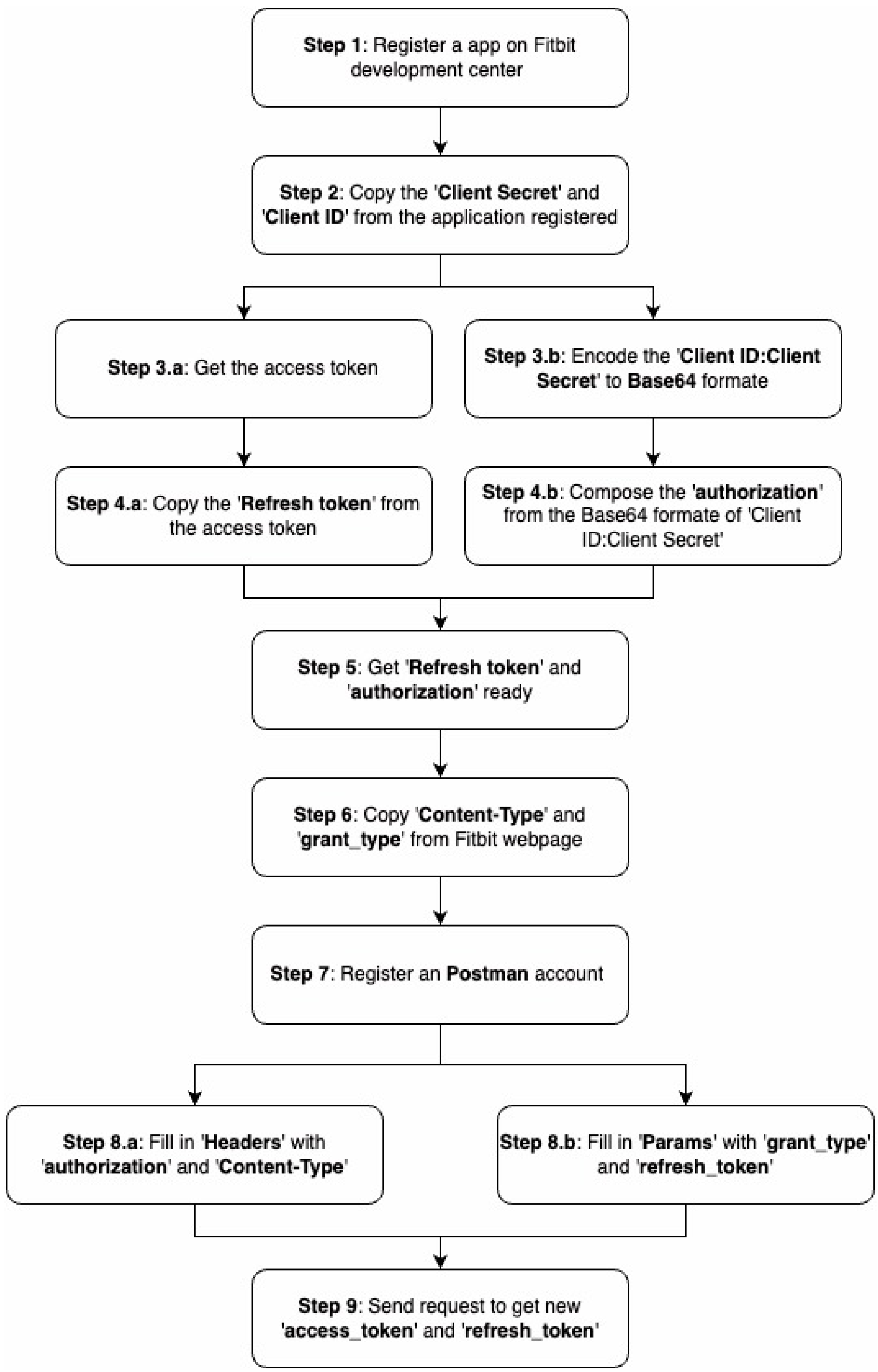
To ensure the security of participants' Fitbit data, we transitioned from the implicit flow to the explicit flow for access tokens. The implicit flow, which exposes the access token to the client application, is considered less secure. In the explicit flow, also known as the authorization code flow, the client application redirects the user to the authorization server for authentication and permission granting. After successful authentication and authorization, the authorization server returns an authorization code to the client application. The client application then exchanges this code for an explicit access token by making a request to the authorization server's token endpoint. The client includes the authorization code and its client credentials (client ID and client secret) in the token request. The authorization server verifies the code and issues an access token if valid. The client application uses this explicit access token to make API requests on behalf of the user, typically including it as an authorization header or in the request parameters.

The explicit flow offers greater security as the access token is not exposed to the client application or visible in the browser's address bar. Additionally, using client credentials in the token request adds an extra layer of protection. The explicit access token flow is commonly used for server-side applications or when accessing sensitive resources in the Fitbit Web API or any other OAuth 2.0-compliant API. Our web application functions as a client application and utilizes the explicit access token flow to access participants' Fitbit data.

5.2.2 Refresh Token

To best protect the security of users' data, the access token has a limited validity period (e.g., 8 hours in this case). When the access token expires, the third-party application (i.e., MissFit) can no longer use it to make requests. To gain continued access, we must send a token refresh request to Fitbit's token endpoint using the refresh token obtained during the initial authorization process. The flowchart in Figure 18 demonstrates the steps of refreshing the access token using Postman. Postman is a popular platform that provides a user-friendly interface for testing APIs. In this case, we use Postman to send a token refresh request to Fitbit for a new access token.

Figure 18: Procedure of refreshing access token via Postman.



5.2.3 JSON Data Stoarge

As shown in Figure 17 (bottom left), the JSON Data Storage initiates a fetch data request to Fitbit (Step 4.a). Upon receiving the requested exercise data, it processes it into a specific format (Step 5.a), stores the processed data in a JSON file (Step 6), hosts the JSON file on a dedicated JSON server (Step 7), and manages data fetching requests from MissFit frontend (i.e., the GUI) (Step 8). This setup allows MissFit to retrieve data directly from the JSON server, effectively bypassing the rate limit imposed on Fitbit API calls.

We used the JSON server as our backend solution. The JSON server is a straightforward, lightweight, and open-source tool designed to rapidly set up a RESTful API, supporting CRUD operations (Create, Read, Update, Delete) tailored for testing and prototyping purposes. While it's commonly used for prototyping and testing, its simplicity and ease of use make it an excellent choice for scenarios where a full-fledged backend isn't necessary. JSON Server allows us to define routes and data structures and perform CRUD operations using a JSON file as a database. This streamlined approach facilitates rapid development and testing of front-end components without the need for a complex backend infrastructure. It provided the necessary API endpoints for our front end to interact with data.

As depicted in Figure 19, the figure showcases three data examples stored in the JSON server. While only three examples are shown for illustrative purposes, it's important to note that MissFit utilizes more data sources than the depicted three. MissFit specifically retains the most recent fourteen days of relevant data, and the data size varies based on users' exercise activity, ranging from 111 KB to 3.6 MB. The data type occupying the most storage is the GPS location data, comprising millions of latitude and longitude pairs that constitute the entire route from location A to location B with good accuracy.

5.3 System Architecture

MissFit is a functional web app designed to assist users in estimating their missing data on Fitbit. The system architecture, as shown in Figure 20, comprises three main components: the

Figure 19: Examples of data sources using the JSON format in MissFit.


```
"data": [
  {
    "activityName": "Run",
    "dateTime": "2022-09-05",
    "distance": 0.319671,
    "duration": 6389000,
    "startTime": "2022-09-05T18:16:41.000-04:00",
    "activeDuration": 6385000
  },
  {
    "activityName": "Run",
    "dateTime": "2022-09-03",
    "distance": 2.35632,
    "duration": 4508000,
    "startTime": "2022-09-03T18:16:33.000-04:00",
    "activeDuration": 4506000
  },
  {
    "activityName": "Run",
    "dateTime": "2022-09-02",
    "distance": 2.400695,
    "duration": 2912000,
    "startTime": "2022-09-02T19:31:45.000-04:00",
    "activeDuration": 2910000
  },
  {
    "activityName": "Run",
    "dateTime": "2022-08-31",
    "distance": 0.747938,
    "duration": 1843000,
    "startTime": "2022-08-31T10:40:42.000-04:00",
    "activeDuration": 1839000
  },
```


(a) Running activities.


```
"data": [
  {
    "dateTime": "Bloomfield",
    "value": 16
  },
  {
    "dateTime": "Friendship",
    "value": 2
  },
  {
    "dateTime": "North Oakland",
    "value": 1
  },
  {
    "dateTime": "Central Oakland",
    "value": 2
  },
  {
    "dateTime": "East Liberty",
    "value": 1
  }
],
```


(b) Neighborhood


```
"data": [
  [
    {
      "lat": 40.45975720882416,
      "lng": -79.93900775909424
    },
    {
      "lat": 40.45975720882416,
      "lng": -79.93900775909424
    },
    {
      "lat": 40.45975720882416,
      "lng": -79.93900775909424
    },
    {
      "lat": 40.45975720882416,
      "lng": -79.93900775909424
    },
    {
      "lat": 40.45975720882416,
      "lng": -79.93900775909424
    },
    {
      "lat": 40.45975720882416,
      "lng": -79.93900775909424
    },
```


(c) GPS

Note: The figure illustrates three examples of data sources stored in the Data Storage Server using the JSON format: (a) Displays four instances of running activities recorded by the user. The data is fetched from Fitbit, and we only extracted relevant information to MissFit. (b) Presents all instances of neighborhoods the user has run or walked in over the past three months. The neighborhood data is obtained through a reverse engineering method from the latitude and longitude in the GPS. To maintain accuracy, we extract the initial 10 seconds of latitude and longitude pairs from the GPS data, average them into one pair, and apply reverse engineering techniques to that pair. (c) Exhibits six instances of GPS data, comprising latitude and longitude pairs fetched from Fitbit. It's worth noting that this data is initially stored in Training Center XML (TCX) format from Fitbit. Each TCX file contains one route, and each fetch request retrieves one TCX file. After obtaining all TCX files, they are processed into the JSON format.

Fitbit Web API, the MissFit Web App, and the Heroku Cloud Service.

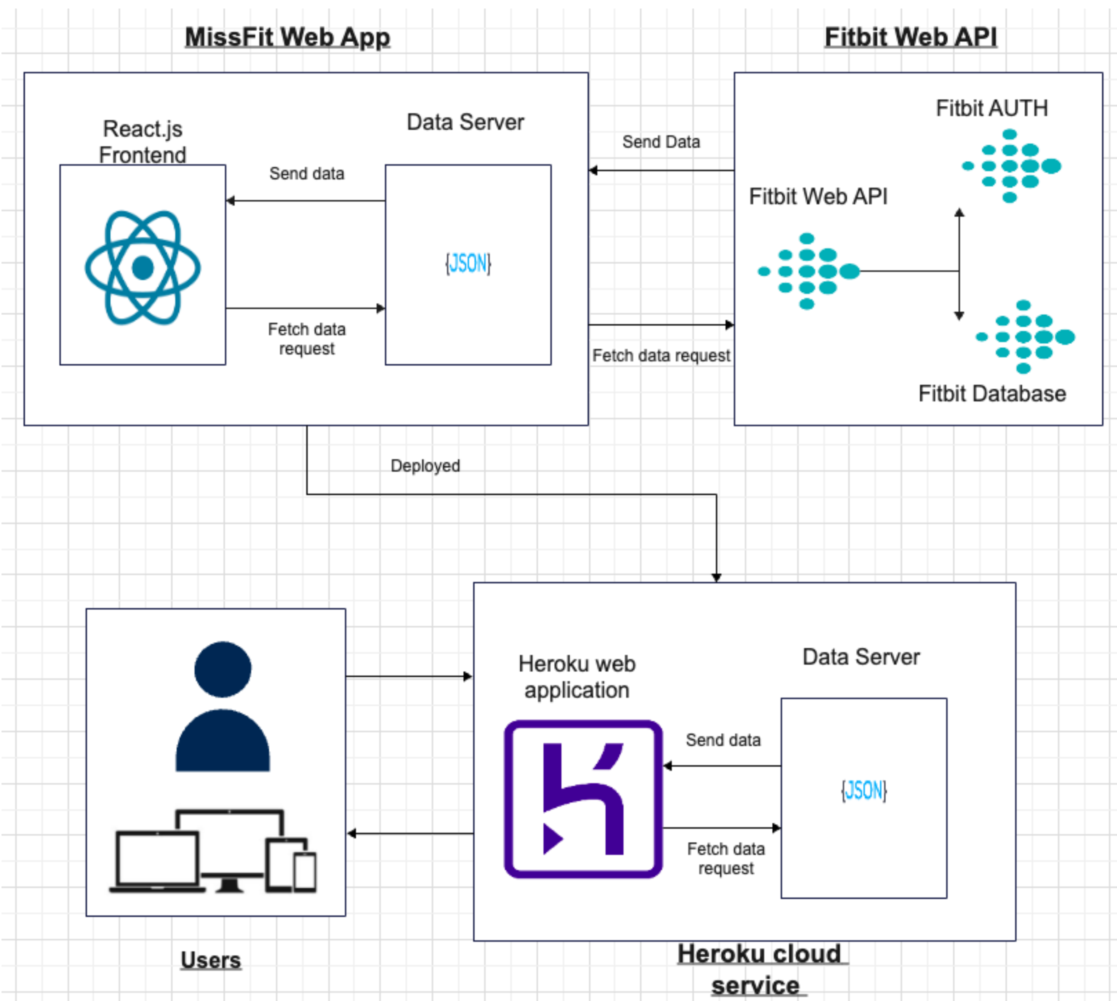
The front end of MissFit is developed using React version 17.0.2, providing a dynamic and responsive user interface. The back end leverages JSON server version 0.17.1 to manage data storage and retrieval efficiently. React is a powerful and popular JavaScript library for building user interfaces. Maintained by Facebook, it is renowned for its declarative and efficient approach to building user interfaces. With its component-based architecture, React allows us to encapsulate and modularize different parts of our application, promoting code reusability and maintainability. The virtual DOM feature optimizes rendering performance, ensuring a smooth and responsive user experience.

The Fitbit Web API is responsible for handling data-fetching requests initiated by MissFit, ensuring seamless communication with Fitbit servers to access user data.

MissFit is deployed on the Heroku cloud service, offering a reliable and scalable hosting solution for web applications. The choice of Heroku is attributed to its ease of deployment, scalability features, and minimum effort for maintenance.

This architecture allows MissFit to deliver a user-friendly experience while effectively estimating missing data by interfacing with the Fitbit Web API and leveraging the capabilities of the Heroku Cloud Service for deployment.

Figure 20: System architecture for MissFit.



6.0 Towards the Design of Systems Improving the Ability to Mediate Missing Data

In Section 3, we conducted an investigation into various existing fully computational methods for dealing with missing data. However, we found that algorithmic models struggle to provide personalized estimation for missing data, prompting us to shift towards a more human-centered approach in eliciting missing data in personal tracking systems. In Section 4, we interviewed participants who engaged in personal tracking to understand their experiences with missing data. From this study, we identified two distinct user groups: trainees and maintainers, each with their specific objectives and goals. Trainees are individuals with concrete and, in some cases, professional athletic training goals, relying on data to inform their decisions and improve performance-related metrics. They are primarily interested in specific insights and precise data comparisons to track their progress accurately. On the other hand, maintainers have more general health improvement and maintenance goals, seeking a holistic understanding of their progress through summative insights from past activities. Detailed comparisons or specific data points are less important to maintainers as they focus on the broader picture of their health journey. We decided to focus on maintainers for this study as they represent a more widespread population, allowing us to cater to the needs of a broader user base by designing for maintainers.

To further explore the needs of maintainers, we implemented three mechanisms in MissFit: algorithmic-based, manual input, and event-based (See Section 5 for implementation details). These mechanisms were informed by the implications derived from our previous findings in Section 4 and aimed to assist maintainers in filling in their missing data. The objective was to enable maintainers to understand their past, gain insights into their present, and strike a balance between different tolerance levels of missing data, visual representation of missing data vs. no data at all, and maintaining trust in their data.

In this study, our primary goal is to refine the designs of PI systems to enhance the ability to mediate missing data for individuals who are tracking to maintain their health. We aim to achieve this by using real-world data, gathering more detailed feedback to support their needs better, and examining the role of synthetic data and the implications of various mechanisms. Through these efforts, we strive to improve the design implications and functionalities of PI tools for maintainers, making a valuable contribution to the broader field of personal informatics.

To guide our research efforts, we have formulated the following research questions:

- ***RQ1: Evaluating mechanisms to mediate missing and synthetic data representations in relation to Personal Informatics.***
  1. Are the strategies implemented in our system effective in communicating synthetic data in the context of real data?
  2. How does the elicitation mechanism impact the perception of synthetic data accuracy?
  3. How does the elicitation mechanism impact the perception of trust in the synthetic data?
  4. Does the elicitation mechanism investigation expose fundamental limitations in trust and the use of synthetic data?
- ***RQ2: Evaluating the impact of different mechanisms.***
  1. Do the different elicitation mechanisms impact how users reflect in setting goals?
- ***RQ3: Finding insights into different mechanisms.***
  1. Is there a preference for utility and perceived value for using a specific elicitation mechanism?
  2. Do the specific elicitation mechanisms have limitations or weaknesses regarding specific missing data situations or personal (training) tracking goals?

6.1 Methods

6.1.1 Use of iterative user-centered design

To answer our research questions, we employed the user-centered design [2] methodology. It's an iterative design process in which the needs, preferences, and behaviors of end users are given primary consideration at every stage of product development. The goal is to create intuitive, efficient, and enjoyable products, services, or systems by ensuring they meet the needs and expectations of those who interact with them. An iterative user-centered design study typically involves several key steps:

- User Research: This phase involves gathering information about the target users, including their demographics, preferences, goals, and pain points. Interviews, surveys, and observations are used to gain insights into user needs and behaviors.
- Requirements Gathering: Based on the findings from user research, designers define the requirements and objectives of the project. This step helps ensure that the design team has a clear understanding of what needs to be achieved and what criteria the design must meet to be successful.
- Design Ideation: In this phase, designers brainstorm and generate ideas for how to address the user needs identified in the research phase. Sketches, wireframes, and prototypes may be created to explore different design concepts and solutions.
- Prototyping and Testing: Designers create prototypes or mockups of the product or system to test with users. Usability testing sessions are conducted to gather feedback on the prototype's usability, functionality, and overall user experience. This feedback is used to refine the design and make improvements.
- Iteration: Based on the feedback received from user testing, the design is iterated upon, with further refinements made to address any usability issues or concerns identified by users. This process of testing, feedback, and iteration continues until the design meets the needs and expectations of the users.
- **Implementation**: Once the design has been refined and validated through user testing, it is implemented and launched for use by the target audience. Ongoing monitoring and feedback collection may continue after launch to ensure that the design remains effective and continues to meet user needs over time.

The iterative user-centered design methodology utilized in this study serves two primary purposes. Firstly, it ensures that the design outcome aligns with the needs, preferences, and aspirations of the individuals (referred to as maintainers) who will use the future PI tools. Secondly, it enables us to integrate the ideas and insights of our end users into the design process, fostering a sense of ownership and collaboration.

For instance, in addressing RQ1, the user-centered design methodology enables us to actively engage participants in sketching out their desired data representations if they express dissatisfaction with our initial designs. Moreover, if they identify elements that are not included in our prototype, we can request them to sketch their ideas, thereby incorporating their perspectives into the design process. This process is conducted in multiple sessions, and improvements and refinements are made based on feedback and results obtained from previous iterations (see Section 6.1.3 for more details).

By conducting the study iteratively, we can also identify user needs that were not included in the previous prototype and were not emphasized by our end users. After incorporating our interpretation of users' feedback, we can test our interpretations against the users' ideas and evaluate the satisfaction of the incorporated design.

6.1.2 Participants

We recruited participants through the Pitt+Me research participation program at the University of Pittsburgh. Pitt+Me serves as a platform connecting individuals interested in participating in research studies with researchers conducting various types of studies at the university. Participants can sign up for Pitt+Me, provide their information and preferences, and researchers can match them with relevant studies based on their interests and eligibility criteria.

For our study, we implemented specific inclusion criteria for participant selection. These criteria included the following: 1) having a minimum usage duration of three months with a Fitbit device, 2) engaging in outdoor running, walking, or hiking activities at least twice per week, 3) possessing a Fitbit account to track and reflect on activity data, 4) owning or consistently accessing a Fitbit device, 5) having a goal of maintaining health, and 6) ensuring continued usage of their Fitbit device throughout the duration of the user study.

To reach potential participants, we distributed recruitment posters featuring Pitt+Me advertisements across various locations, including university buildings, gyms, restaurants, and communities. However, all recruitment efforts were coordinated through the Pitt+Me platform. We received 63 referrals from Pitt+Me, and upon receiving these referrals, we conducted phone screenings with the potential participants to determine their eligibility for the study. If deemed eligible, we then scheduled the first appointment with them.

A total of 14 individuals participated in the study, comprising 12 females and 2 males. The participants' ages ranged from 19 to 56 years, with a mean age of 32.6 and a standard deviation

Table 12: Participants' demographics and frequency of exercise from the iterative user-centered design study.

PID	Age range	Gender	Wearables	Frequency of exercise	Employeement status
P1	40-49	Female	Fitbit	1-2 times per week	Employed, working 1-20 hours per week
P2	30-39	Female	Fitbit	5-6 times per week	Employed, working 21-40 hours per week
P3	21-29	Female	Fitbit	4-5 times per week	Employed, working 21-40 hours per week
P4	50-59	Female	Fitbit	4-5 times per week	Employed, working 21-40 hours per week
P5	21-29	Female	Fitbit	5-6 times per week	Student
P6	21-29	Female	Fitbit	1-2 times per week	Employed, working 21-40 hours per week
P7	40-49	Male	Fitbit	1-2 times per week	Employed, working 41 or more hours per week
P8	18-20	Female	Fitbit	4-5 times per week	Student
P9	30-39	Female	Fitbit	5-6 times per week	Student
P10	21-29	Female	Fitbit	5-6 times per week	Employed, working 41 or more hours per week
P11	21-29	Male	Fitbit	Above 7 times per week	Employed, working 21-40 hours per week
P12	40-49	Female	Fitbit	Above 7 times per week	Employed, working 41 or more hours per week
P13	21-29	Female	Fitbit	4-5 times per week	Student
P14	30-39	Female	Fitbit	3-4 times per week	Employed, working 21-40 hours per week

of 10.59. All participants were residents of the Pittsburgh area. Among them, 21.4% reported engaging in walking, running, or hiking activities one to two times per week; 14.3% reported doing exercises three to four times per week; 28.6% of the participants reported doing exercises four to five times per week; 28.6% reported doing exercises five to six times per week; and 14.3% reported doing exercises more than seven times per week (See Table 12 for details).

To get a broader context for understanding participants and their responses and potentially enhance the study's validity, generalizability, and overall impact, we additionally gathered self-reported wellness data from each participant, utilizing a questionnaire adopted from [78], the questionnaire collected information under the guidance of the Wellness Wheel concept. The Wellness Wheel concept encompasses seven significant dimensions of wellness: emotional, environmental, intellectual, occupational, physical, social, and spiritual [77]. It acknowledges that well-being is multifaceted and that each dimension contributes to an individual's overall quality of life [72]. We employed this questionnaire to gain a foundational understanding of the wellness status of our participants. The findings reveal that all participants scored in the outstanding range across all seven wellness spectrums, affirming that our participants are individuals who prioritize and dedicate efforts toward their overall well-being.

6.1.3 Procedure

The study consisted of four visits, with the first visit offering the option of being conducted virtually or in person, and the subsequent three visits being conducted in person. The iterative user-centered design approach was implemented during the second, third, and fourth visits. Each participant was scheduled to attend one visit per week, spanning a duration of one month, with the possibility of extending the duration if there was insufficient recorded activity or if personal circumstances required participants to leave town.

6.1.3.1 Visit 1

Once the interview is scheduled, we will provide participants with the consent form via DocuSign. DocuSign is an electronic signature platform allowing users to sign and send documents digitally. It provides a secure and legally binding way to sign documents without needing physical copies or in-person meetings. They are advised to review the form beforehand and address any questions during the first visit. If no questions arise, they can proceed to sign the form. Additionally, participants will receive calendar invitations with the Zoom meeting ID for the virtual session.

During the Zoom meeting, we start by expressing our gratitude to the participants for their willingness to participate in our study. We then proceeded to explain the goals, processes of the study and an overview of the consent form. We invited the participants to ask any questions they may have regarding the study's objectives, procedures, and consent forms. If there were any inquiries, we promptly addressed them.

Following this, the visit continued with a brief introduction and a discussion to gain a general understanding of the participants' Fitbit device usage. Prior to commencing the discussion, we sought their oral consent to record the session for documentation purposes. To gain insight into their Fitbit usage patterns, we posed a series of questions, such as the duration of their Fitbit device usage, their motivations for using the device, the frequency of their outdoor activities while wearing the Fitbit device, whether these activities occurred consistently at the same time or varied, the regularity of their engagement in these activities, the locations where they have previously engaged in running, hiking, or walking over the past six months, their goals pertaining to running, hiking, or walking, the frequency of missing data in their Fitbit records during these activities, and their strategies for dealing with missing data in their Fitbit device. Depending on the depth of the conversation and participant responses, we also asked follow-up questions. These included inquiries about the most recent activities recorded on their Fitbit application and the types of activities they frequently engage in using their Fitbit device(s).

Next, we requested participants to complete a survey regarding their general health conditions [78] and demographics [93]. The survey was distributed using the Qualtrics online survey platform. Qualtrics is a popular and versatile tool that allows researchers and organizations to create, distribute, and analyze surveys and collect participant feedback. To ensure participant privacy, we generated a unique eight-digit participant ID to anonymize their identity.

Finally, we request participants grant us access to their Fitbit activity data using the Fitbit Web API. The Fitbit Web API is an interface Fitbit provides for developers to retrieve and interact with data from Fitbit devices and services. It allows access to various information from Fitbit user accounts, including activity data, sleep data, heart rate, and user profile details. Our web application, MissFit, which utilizes the Fitbit Web API, demonstrates three methods to collect missing data. We require an access token for each participant to access the participants' Fitbit data. An access token serves as a credential that authenticates and authorizes access to protected resources in the API or web application. In the Fitbit Web API context, an access token is generated when a user grants permission to a third-party application to access their Fitbit data. This access token is then used in subsequent API requests to validate the identity and permissions of the requesting application. The technical details of how to get the access token are explained in Section 5.2.1. To safeguard the access token, we store it in an Excel file located in a monitored OneDrive folder managed by the University of Pittsburgh. Each access token expires in eight hours, and we utilize Postman to refresh the access token when necessary. Postman is a widely used collaboration platform and API development tool that simplifies the process of building, testing, and documenting APIs. It supports various authentication methods, making it suitable for testing and interacting with APIs that require authorization. Once participants complete the user study or drop out, we promptly delete their access and refresh tokens from the Excel file to ensure data security.

6.1.3.2 Visit 2

During the second visit, we adhered to the participatory design protocol. We expressed our gratitude for their participation and dedication to the study. We then explained the purpose and process of the visit, highlighting that the prototype presented was in its early stages and subject to changes throughout the iterative design process. We encouraged participants to provide genuine feedback to aid in improving the design. To ensure they understood the focus of the study, we clarified that we were testing the prototype itself and not evaluating their performance, relieving any pressure they may feel about participating.

Before delving into the initial prototype, we discussed the general usage of their Fitbit device(s). We inquired about any changes in their goals for using the Fitbit device(s) and their exercise routines, including frequency, type, locations, and exercise time. Follow-up questions explored any changes in their work or daily routines and how these might impact their exercise routines and perceptions of different methods for that week.

Next, we presented two sets of interfaces with slight modifications, displaying the participants' own Fitbit activity data to make the interfaces more relevant to them. In the first set, we intentionally removed data for an entire day, asking participants to imagine a scenario where they forgot their Fitbit at home. We informed them that data for the day still existed, but we had taken it out for study purposes. With this hypothetical scenario in mind, we asked them to use the three methods to elicit the missing data, requesting them to think aloud while using the methods. After each method, we conducted a short interview and provided a quantifiable scale survey for them to fill in.

For the second set, the data is missing for one specific activity. Here, we ask participants to imagine a scenario where their Fitbit ran out of battery while they were on their walking activity. Similar to the first set, we present the two hypothetical scenarios and proceed with the three elicitation methods. It's important to note that the order of the two different scenarios and the order of the three elicitation methods are randomized to avoid any potential bias in the results. During these scenarios, we encourage participants to think aloud and provide feedback on their experience with each method. After completing the methods, we conduct a short interview and offer a quantifiable scale survey for them to fill in. This allows us to gather their insights and preferences for each elicitation method in both scenarios.

After the participants had interacted with all three methods under the two different scenarios, we conducted a reflective session to delve deeper into their experiences. We engaged in an in-depth discussion on some of the questions asked during the interaction. Additionally, we asked them to provide feedback on what they liked and disliked about each elicitation method. Furthermore, we asked them to rank the designs based on their level of trust in the methods and rank the designs based on their preferences. This feedback and ranking will help us understand their perceptions and preferences towards the different methods and inform our decision-making process for future prototype iterations.

Finally, we engaged in a co-design session with the participants, utilizing pen and white paper. During this session, we revisited the visualization of their activity data and sought their input on possible recommendations or alternate ways they would prefer to represent their data. The researcher carefully documented incidents that occurred during the interview, which included participants expressing experiences not being adequately represented in the initial prototype or sharing their unique approaches to estimating missing data differently from what was presented. In such instances, we encouraged the participants to sketch out their ideas on the white paper, such as recalling the routes they take for their runs or walks and checking their calendars or sleep times to recall the days. This collaborative approach allowed us to gather valuable insights and co-create design ideas to improve the prototype and tailor it to the participant's specific needs and preferences.

6.1.3.3 Visit 3

Before Visit 3, we periodically checked the participants' Fitbit activities to ensure that enough activities were captured for the interview. If the Fitbit data showed a lack of activities for Visit 3, we called the participants to reschedule the interview for the next week and reminded them to collect enough activities. In the meantime, the researcher analyzed the incidents noted and the co-design ideas gathered from the previous iteration, then made further improvements to the prototype based on our own interpretations of the feedback and sketches on the paper from the participants. For example, if a participant expressed a preference for checking cumulative step counts instead of step counts for specific activities, we visualized the cumulative step count for this iteration instead. If a participant expressed using calories to measure their activities, we then utilized calories in the process of estimation. The prototype may combine digital and paper prototypes for faster production of ideas. The focus of the discussion will be on the participants' feedback on the process of the methods as well as testing out our interpretation from the previous iteration. The other processes will remain the same as in Visit 2. We continued to engage with the participants and gather their insights to refine the prototype and make it meet their needs and be more effective.

6.1.3.4 Visit 4

In Visit 4, we followed the same process as in Visit 3. Before the interview, a researcher checked whether the participant had enough activities recorded. If they did, the participants interacted with us using the improved prototype based on interpretations, provided feedback on the methods, and engaged in co-design sessions to further refine the design. Once the interview was completed, we expressed our gratitude to the participants for their participation and compensated them for their time and effort.

6.1.4 Data preparation and analysis

This study had four visits. The first visit included interviews about participants' general usage; the second, third, and fourth visits involved interviews about MissFit, participants' recommendations for modifying MissFit, and notes taken by a researcher during each visit. Thus, the data from this study consisted of audio recordings and notes. The audio files from this study were imported and machine-transcribed using the cloud-based platform Otter.ai[1].

Employing open coding, the researchers identified themes and patterns in the transcripts and explored connections among these codes through axial coding, as outlined by Scott [88]. Initially, the two researchers individually examined and coded 10% of the transcripts, utilizing the previously identified themes. Their initial level of agreement, expressed as inter-rater reliability, was 0.769, surpassing the expected chance agreement of 0.56, as proposed by Krippendorff ([45], p. 224-226). To address any discrepancies and update the existing codebook, the two researchers en-

[1]https://otter.ai/

gaged in discussions. Later, the first researcher completed the coding of the remaining transcripts using NVivo, a software program designed for qualitative and mixed-methods research. NVivo is widely utilized by researchers, particularly in the social sciences and other fields where qualitative data analysis is common [103].

The relationships of the interview data with the results section (see Section 6.2 for more details) are as follows:

- The four dimensions are themes identified using the interview transcripts from the first visit, where participants express their motivation for using Fitbit and how they integrate it into their daily lives.
- The design observations were identified using interviews from the second, third, and fourth iterations, where participants provided feedback after using each method provided in MissFit.

6.2 Results

After analyzing the interview data, we observed that our participants can be categorized into four distinct groups based on two dimensions: ***motivation*** and ***exercise routines***. Further examination of the interview data revealed that motivations from participants can be categorized into *wellness-oriented* and *performance-oriented* based on their personalities. At the same time, exercise routines can be categorized as either *structured* or *unstructured*.

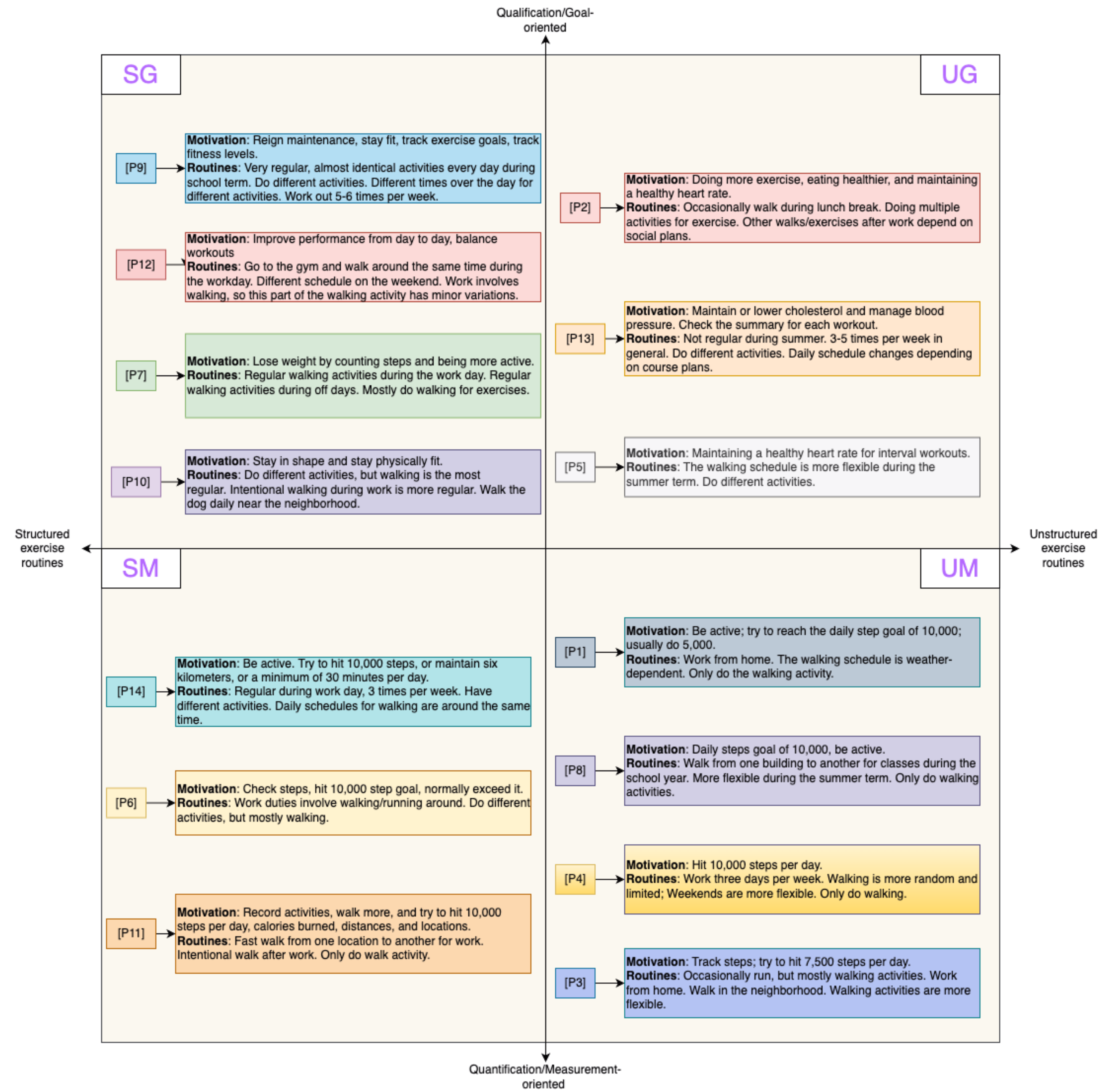
We define wellness-oriented fitness goals as objectives centered around achieving broad but definable outcomes related to health, fitness, or overall well-being. Individuals with personalities dominated by this approach emphasize their desired outcomes or achievements without necessarily linking them to specific numerical values. For example, a participant may aim to experience improvements in their cardiac performance through aerobic or cardiovascular exercises such as running, cycling, or brisk walking. Even though they may monitor numerical values for these activities, their higher-level goal is to enhance performance, and their desired outcome is often explicit but challenging to measure with a specific numerical value. On the other hand, we define performance-oriented fitness goals as fitness objectives that are well-defined and monitored using measurable parameters. Individuals with personalities dominated by this approach often set goals that are clear, precise, and can be assessed using quantifiable metrics. For instance, a participant may aim to achieve 10,000 steps per day, exercise three times per week, or walk for 30 minutes per day. Their desired outcomes or improvements are often explicit and can be measured with specific numerical values.

In addition, a structured exercise routine is defined as a planned and organized sequence of physical activities throughout the week and day. For example, a participant may plan to run in the gym at 5 a.m., briskly walk at noon, and engage in a slow-paced walk at 7 p.m. Some individuals may follow such a plan every day from Monday to Friday, while others might choose to follow a specific plan three days a week and a different one on weekends. We categorize participants who predominantly follow structured exercise routines as those who plan their weekly and daily workouts at least 70 percent of the time, with minor variations in their schedules. Conversely, an unstructured exercise routine refers to a more flexible and spontaneous approach to physical activity, where individuals engage in exercises or activities without adhering to a predefined plan or schedule. For instance, in a given week, one might briskly walk on three consecutive days at different times, but engage in a slow-paced walk on random days the following week. Participants who are dominated by unstructured exercise routines are those whose exercise patterns do not follow a specific schedule at least 70 percent of the time, even though they may follow a certain structure for short periods. In most cases, their exercise routines are more random in nature.

Based on the four dimensions described above, we have classified our participants into four groups, as depicted in Figure 21. It is important to emphasize that these four groups represent a continuum and are not exclusive; rather, the reader should perceive them as shades on a spectrum. Analogous to personality traits, they are not simple binary categories. These groups provide a framework for understanding participants' motivations and behaviors throughout this study. They proved useful, as we observed shifts in users' feedback across different positions on the spectrum based on the solutions and evolutions from the results.

- ***Structured Wellness-oriented (SW)***: These individuals adhere to a structured exercise routine and exhibit a wellness-oriented personality. For example, consider the case of P12 within this group, as illustrated in Figure 22. Her higher-level goal is to improve her performance in the activities. She consistently runs at the gym at 5 a.m. before starting work, engages in a brisk walk at noon, and follows it up with another brisk walk from work to home on her workdays.

Figure 21: Participants' group information.



Note: The figure presented illustrates the group information of our participants from the iterative user-centered design study, divided by their *motivations* and *exercise routines*.

On her rest days, she shifts her running routine to 9 a.m. and includes more intense brisk walks, albeit with a different schedule than her workdays. Additionally, she occasionally incorporates aerobic workouts into her routine. Notably, her walking activities between her schedules are primarily attributed to her demanding work duties, which have minimal variations.

- ***Structured Performance-oriented (SP)***: Members of this group also follow a structured exercise routine, but they are characterized by a performance-oriented personality. For instance, P14, as illustrated in Figure 23. Her main goal is to achieve 10,000 steps per day and be more active, but she typically manages around 5,000 steps. To reach her target, she plans brisk walks after dinner around her neighborhood, normally around 8 p.m., but sometimes it's 9 p.m. Additionally, she allocates time for swimming at the pool on Saturdays. Due to her sedentary lifestyle, she tries to do one activity each day to stay active.
- ***Unstructured Wellness-oriented (UW)***: This category includes participants who engage in an unstructured exercise routine and are primarily wellness-oriented. For example, as shown in Figure 23, P5's primary objective is to monitor her heart rate performance during interval workouts. Throughout the day, her exercise routine varies. She swims at the pool, occasionally incorporates brisk walks between home and school, and participates in frisbee tryouts during the summer term. Her schedule is not only flexible from day to day but also varies throughout the day. Another example in this group is P13, as shown in Figure 25, whose primary objective is to manage blood pressure. Her schedule is also quite flexible. She sometimes engages in sports or runs three days in a row, but other times, she opts for brisk walks, and occasionally she has days with no exercises.
- ***Unstructured Performance-oriented (UP)***: This category includes participants who engage in an unstructured exercise routine and are primarily performance-oriented. Consider the example of P3, as shown in Figure 26. Her primary objective is to achieve 7,500 steps per day and maintain an active lifestyle. Working from home provides her flexibility, allowing her to take her dog for walks in her neighborhood three times per week. However, her walks vary in terms of timing and frequency, reflecting the unstructured nature of her exercise routine.

Figure 22: Participants' exercise routine.

	Thu 07/13	Fri 07/14	Sat 07/15	Sun 07/16	Mon 07/17	Tue 07/18	Wed 07/19	Thu 07/20	Fri 07/21	Sat 07/22	Sun 07/23	Mon 07/24	Tue 07/25
3:00 AM													
4:00 AM	Walk												
5:00 AM	Run	Run			Run	Run			Run			Run	Run
6:00 AM													
7:00 AM													
8:00 AM				Walk									
9:00 AM				Run					Walk		Run		
10:00 AM			AW	Walk			Walk			Walk	Walk		
11:00 AM				Walk	Walk						Walk		
12:00 PM		Walk	Walk				Walk	Walk		Walk	Walk		
1:00 PM	Walk	Walk	Walk		Walk	Walk		Walk	Walk			Walk	Walk
2:00 PM	Walk			Walk	Walk	Walk		Walk	Walk	Walk		Walk	
3:00 PM								Walk					
4:00 PM						Walk	Walk			Walk	Walk	Walk	
5:00 PM	Walk	Walk			Walk				Walk	Walk		Walk	Walk
6:00 PM		Walk		Walk			Walk	Walk		Walk			
7:00 PM			Walk	Walk			Walk						
8:00 PM		Walk							AW				
9:00 PM		Walk								Walk			
10:00 PM													

Note: Above is an example of P12's weekly and daily exercise routine in group SW, categorized as follows based on records from her Fitbit: 'Walk' for brisk walking outside, 'Run' for running at the gym, and 'AW' for aerobic workouts.

Figure 23: Participants' exercise routine.

	Tue 5/30	Wed 5/31	Thu 6/01	Fri 6/02	Sat 6/03	Sun 6/04	Mon 6/05	Tue 6/06	Wed 6/07	Thu 6/08	Fri 6/09	Sat 6/10	Sun 6/11	Mon 6/12
8:00 AM														
9:00 AM									Walk					
10:00 AM														
11:00 AM														
12:00 PM														
1:00 PM					Swim									
2:00 PM														
3:00 PM														
4:00 PM												Swim		
5:00 PM														
6:00 PM														
7:00 PM	Walk													
8:00 PM				Walk		Walk				Walk				
9:00 PM		Walk						Walk						Treadmill
10:00 PM			Walk				Walk							

Note: Above is an example of P14's weekly and daily exercise routine in group SP, categorized as follows based on records from her Fitbit: 'Walk' for brisk walking outside, 'Swim' for swimming at a pool, 'Treadmill' for running at the gym.

Figure 24: Participants' exercise routine.

	Wed 5/03	Thu 5/04	Fri 5/05	Sat 5/06	Sun 5/07	Mon 5/08	Tue 5/09	Wed 5/10	Thu 5/11	Fri 5/12	Sat 5/13	Sun 5/14	Mon 5/15
8:00 AM													
9:00 AM												Sport	
10:00 AM				Sport	Sport					Walk		A W	
11:00 AM				A W	A W							Sport	
12:00 PM				Sport	Sport			Walk			A W		
1:00 PM				A W							Sport		
2:00 PM											Sport		
3:00 PM					Sport						A W		
4:00 PM	Swim												
5:00 PM													
6:00 PM						Sport		Walk					Sport
7:00 PM						A W			A W				
8:00 PM		Walk				A W							
9:00 PM													

Note: Above is an example of P5's weekly and daily exercise routine in group UW, categorized as follows based on records from her Fitbit: 'Walk' for brisk walking outside, 'Sport' for frisbee, and 'AW' for aerobic workouts, 'Swim' for swimming in the pool.

Figure 25: Participants' exercise routine.

	Thu 6/01	Fri 6/02	Sat 6/03	Sun 6/04	Mon 6/05	Tue 6/06	Wed 6/07	Thu 6/08	Fri 6/09	Sat 6/10	Sun 6/11	Mon 6/12	Tue 6/13
10:00 AM													
11:00 AM				Sport									
12:00 PM		Walk		Walk									
1:00 PM		Weights			Sport					Run			
2:00 PM				Walk		Sport							
3:00 PM													
4:00 PM													
5:00 PM								Walk					
6:00 PM					Weights			AW					
7:00 PM													
8:00 PM								Run					
9:00 PM		Walk											
10:00 PM													
11:00 PM		Walk							Run				
12:00 AM													

Note: Above is an example of P13's weekly and daily exercise routine in group UW, categorized as follows based on records from her Fitbit: 'Walk' for brisk walking outside, 'Weight' for weight lifting, 'AW' for aerobic workouts, 'Run' for running around with kids due to her babysitting job, and 'Sport' for playing with kids due to her babysitting job.

Figure 26: Participants' exercise routine.

	Mon 5/08	Tue 5/09	Wed 5/10	Thu 5/11	Fri 5/12	Sat 5/13	Sun 5/14	Mon 5/15	Tue 5/16	Wed 5/17	Thu 5/18	Fri 5/19	Sat 5/20
8:00 AM													
9:00 AM													
10:00 AM				walk								walk	
11:00 AM					walk			walk					
12:00 PM								walk					
1:00 PM				walk				walk					
2:00 PM											walk	walk	
3:00 PM	walk				walk								
4:00 PM													
5:00 PM				walk									
6:00 PM		walk					Walk						
7:00 PM				walk									
8:00 PM						walk	walk						
9:00 PM													

Note: Above is an example of P3's weekly and daily exercise routine in group UP, categorized as follows based on records from her Fitbit: 'Walk' for brisk walking outside with her dog.

6.2.1 Group SW

As depicted in the upper left quadrant of Figure 21, participants in this group are motivated to achieve various wellness goals, including weight maintenance, fitness level tracking, overall fitness maintenance, exercise goal tracking, daily performance comparisons, continuous improvement, and balancing their workouts between gym-based weight-lifting and brisk walking outdoors. Their specific objectives encompass weight loss through increased physical activity, as well as performance tracking for different workout types, such as cardio workouts. Below is a persona representing the typical characteristics of this group observed in our study.

Mark is a 35-year-old IT project manager with a deep passion for data analysis. He thrives on structure and believes in the significance of long-term and higher-level goals as the key to achieving success. His primary fitness objective is maintaining a healthy weight and embracing a healthier lifestyle. Mark relies on his Fitbit device to track his progress and ensure continuous improvement. He diligently tries to workout every day, believing that this practice is instrumental in reaching his ultimate fitness goal. Mark's planned exercise routine unfolds as follows: From Monday to Friday, he dedicates 30 minutes to treadmill running in the morning, followed by a brisk 30-minute walk during his lunch break and another 30-minute fast-paced walk in the evening. When the weekend arrives, his routine shifts slightly to include a 20-minute treadmill run in the morning, followed by 30 minutes of cycling and concluding with another 30-minute brisk walk. Occasionally, he adds some excitement to his weekends by going hiking and other exercises with friends. Mark takes the data recorded on his Fitbit very seriously. Therefore, it can be quite frustrating for him when some of his workout data goes missing due to incidents like forgetting to wear his Fitbit after charging it or signal loss. For example, a missing record of a 30-minute brisk walk will leave a blank in his weekly schedule and ruin his record, which prevents the accurate capture of his workout information. With his in-depth knowledge of his fitness performance and a keen eye for daily progress, Mark places a high premium on data accuracy. He seeks sophisticated and explainable solutions and support to regain his missing data.

6.2.2 Group SP

Participants in group SP are those who have a structured exercise routine and a strong tendency toward precise numerical measurements. As shown in the lower left quadrant of Figure 21, these participants are driven by specific fitness metrics. They are motivated to monitor their daily step count, primarily focusing on achieving or surpassing a specific daily step count (i.e., 10,000) benchmark. Their objectives include tracking their activity levels and consistently striving to achieve and exceed a specific step target. Additionally, they closely monitor metrics such as calories burned, distance covered, and exercise locations to indicate whether they've been active or not. Below is a persona representing the typical characteristics of this group observed in our study.

Sarah is a 28-year-old marketing manager who leads a busy and predominantly sedentary lifestyle due to the nature of her job. She recognizes the importance of staying active to counter the effects of prolonged sitting but prefers a flexible approach to her fitness routine. Sarah's primary fitness goal is to reach a daily step count of 10,000 to indicate that she's active enough for the day. She uses Fitbit to keep herself accountable and motivated. Her Fitbit also records the distance and calories she burned for each activity, but she cares most about whether she reached her daily steps or not. She glances at other metrics Fitbit provides but rarely remembers them. Sarah's structured exercise routine is planned as follows: Monday through Friday (3 times), 30 minutes of slow-paced walking during her lunch break, 30 minutes of fast-paced walking in the evening, and 10 minutes of yoga for flexibility before bed. On weekends, she continues with 30 minutes of slow-paced walking in the evening and occasionally engages in sports and other exercises with her friends. Besides that, she also climbed the slopes as much as possible. Sarah doesn't care too much about the exact number of steps she takes every day, but she likes to see that she accomplishes her step goal and receives rewards for completing it. Therefore, she finds it frustrating when the record shows that some of her steps are below her target due to missing data when she forgets to wear her Fitbit. She's not particularly interested in remembering every detail of the metrics Fitbit provides; she's mainly focused on knowing whether she reached her daily step goal or not. Due to her desire for simplicity and her limited interest in memorizing various Fitbit metrics, Sarah seeks solutions and methods that can effectively represent her daily activity without requiring significant effort on her part.

6.2.3 Group UW

As illustrated in the upper right quadrant of Figure 21, these participants are motivated to attain a range of wellness goals. These goals include healthier eating habits, increased physical activity, healthy heart rate during interval workouts, and tracking sleep patterns. Additionally, they are focused on cholesterol management and blood pressure control. The following persona exemplifies the common traits observed within this group in our study.

Mike is a dedicated Ph.D. student in his third year of psychology studies. Having completed all his required classes, he is now fully focused on his thesis research. Throughout his demanding academic journey, it was challenging for Mike to maintain a healthy lifestyle. However, as he nears the finish line, he is determined to prioritize his health once more. To achieve this, Mike has decided to exercise more than before and maintain a healthy heart rate during physical activities. He relies on his Fitbit device to monitor his workouts and periodically assess his progress. Given the demanding nature of his studies, Mike can't dedicate much daily attention to his exercise data. Instead, he looks to his Fitbit to provide a comprehensive overview of his progress and confirm that his exercise promotes heart health. Mike tries to incorporate his exercise routine into his daily life seamlessly. He aims to make it to school three times a week, incorporating brisk walking into his commute between home and school. He also enjoys changing his work environment by occasionally visiting coffee shops and libraries, where he continues to incorporate brisk walking. On weekends, he runs in his neighborhood and occasionally engages in activities like hiking with friends. There are days when he has no activities recorded, and due to the gaps in his exercise data, he sometimes struggles to distinguish between days when he didn't wear his device and days when he didn't engage in exercise. While Mike isn't overly fixated on retrieving every missing data point, he does aspire to achieve a certain level of accuracy in capturing his exercise efforts. He seeks solutions and support to recover the missing data that can reflect his improvements.

6.2.4 Group UP

Participants in group UP maintain an unstructured exercise routine and exhibit a performance-oriented approach in their tracking behavior. As shown in the bottom right quadrant of Figure 21, these individuals are driven by the motivation to monitor their activity levels, such as achieving a

20-minute walk or reaching the goal of 10,000 steps. Their primary focus is on achieving specific step counts or exercise durations rather than recording the specific activities they engage in.

Emily, a 45-year-old housewife, works part-time from home while caring for her two children. Recently, she began experiencing back pain due to her sedentary lifestyle, prompting her to take action and find ways to stay active throughout her busy day. Her primary health goal is simple: to stay active. Although she aims to walk 10,000 steps per day, she often falls short, typically reaching around 5,000 steps. Her exercise routines are sporadic and subject to her daily schedule, which can be unpredictable due to her part-time work and childcare responsibilities. On workdays, she picks up her kids and occasionally enjoys some playtime with them at the park. On weekends, she sometimes goes hiking with her family and takes leisurely walks around the neighborhood with her friends. Emily admits that she's not particularly tech-savvy and sometimes finds it challenging to fully understand and utilize devices like her Fitbit to their full potential. She participates in an initiative linked to her husband's company, where meeting certain activity levels can lead to financial rewards. In support of this program, Emily uses her Fitbit to motivated herself and to record her activities and shares this data with her husband's company. Therefore, it can be frustrating for her when some of her activities go uncounted due to occasional forgetfulness in wearing her Fitbit. She seeks solutions and methods that are easy to understand and accurately represent her daily activity without requiring significant effort on her part.

6.2.5 Design observations

6.2.5.1 Perceptions of reliability and accuracy disparities in scenario context

We presented the prototype we implemented (see Section 5 for more details) to participants. Recall that in scenario A, we presented a situation where some data from one day was missing, described as when their Fitbit was out of battery in the middle of the day. Scenario B, where an entire day was missing, was described as when the user forgot to wear their Fitbit.

In response to the above-described two scenarios, some participants (4 out of 14) expressed that they found the estimated data to be more reliable when some data was captured. This preference was particularly evident in scenario A, where the Fitbit ran out of battery in the middle of the day. For instance, P12 stated, *"it had something captured, to me that's telling me there's something* *that's definitely accurate."* P9 expressed a similar sentiment, saying, *"it is more reliable because there is more data already on that day existing."* P6's response aligned with these views: *"I would trust this one probably more than the original one (missing one day) since there was some data."* These reactions correspond to the persona described in Section 6.2.1, where people are skeptical about the estimated data.

Conversely, some participants (3 out of 14) found the estimation for scenarios involving an entire day missing more accurate. P8 articulated this perspective, stating, *"I just feel like I'd rather simulate the entire day than half the day and add it on."* P9 stated, *"I think the one (missing one day) was much simpler."* Furthermore, some participants believed (7 out of 15) that the estimation for these two scenarios was the same. For example, P11 said, *"What's the difference between these two designs? They look similar."*

6.2.5.2 Varied responses to the algorithmic-based approach

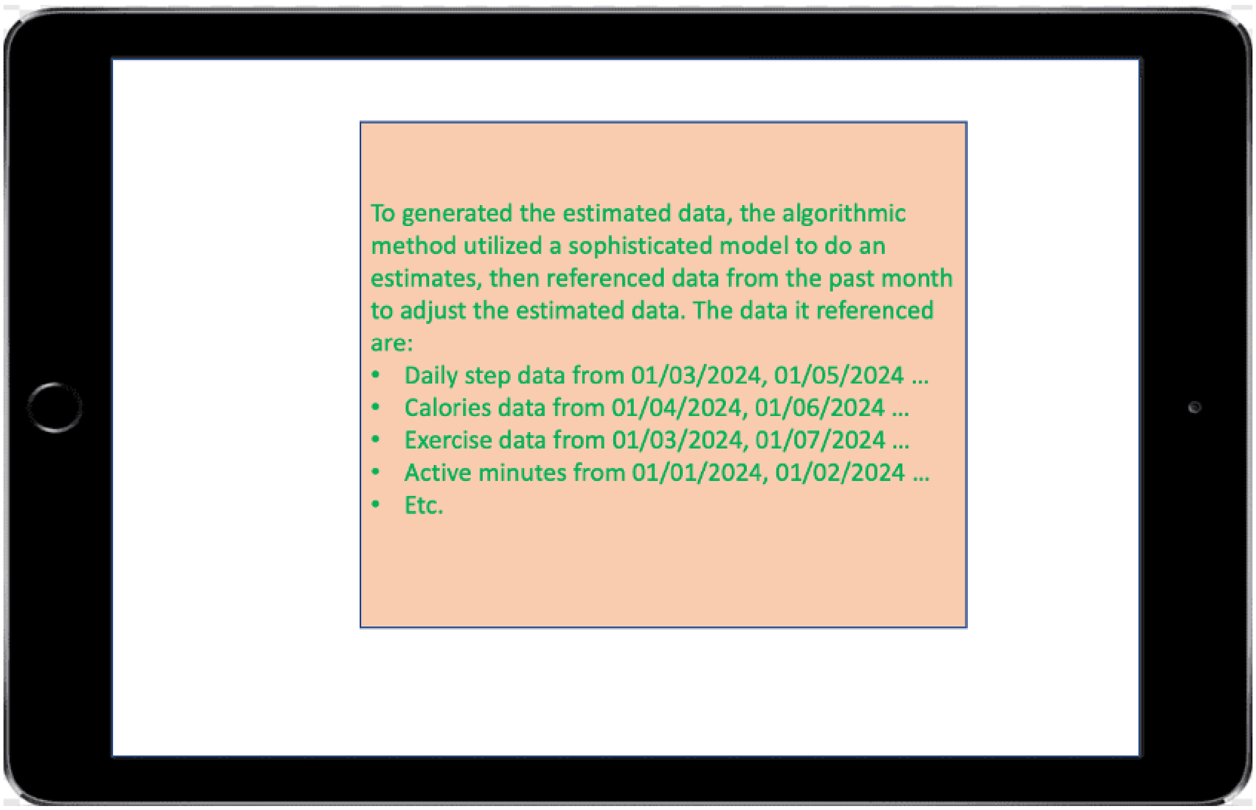
During Visit 2 (see Section 6.1.3.2), while most participants liked the algorithmic-based approach, concerns were also raised. After presenting it, some participants considered the estimated data accurate, but they also desired an understanding of how the algorithm generated these estimates. For instance, P7 mentioned, *"that's (the algorithmic approach) definitely more accurate, but I don't know; the algorithm is boring to me. How is it generating this? What is it basing that on?"* Similarly, P9 commented, *"it's good that I can see that the data has been regained; however, I don't know where that data is coming from."* P3 added, *"I'm not sure how they came up with that number."* However, some participants were more skeptical about the estimation, especially when the estimated data indicated the highest activity level of the week (when we intentionally removed the highest data). For example, P14 expressed uncertainty, stating, *"I feel there is an error, and I'm not really sure."* P8 mentioned, *"if I didn't know exactly what it was doing and how it was calculating it, it feels like an outlier."* P4 questioned, *"I think the estimated data is wrong. How did it generate this?"* Despite expressing a preference for using the approach, some participants conveyed distrust, suggesting that explaining the algorithm's process would foster more trust.

Following the analysis of Visit 2, we identified common patterns with the algorithmic-based approach. Participants favored its ease, minimal effort requirement, and time efficiency. However, it became apparent that there was a need for more clarity on how the system generated the estimated results. And how to improve their confidence in the estimated data, especially for those who only have a vague impression of their step counts. To address this, we had several explorations. Firstly, we added a summary sentence explaining how the algorithm works (see Figure 27 as an example). After presenting it to them, some participants found it helpful in understanding the algorithm, as expressed by P3: *"I like how it explains how the algorithm calculated this by pointing out the data it looked at."* However, some participants reported that, while they care about how the algorithm generated the estimated data, they might ignore lengthy explanations, as noted by P5: *"even if it gave me a long thing telling me how it generated the data, I probably wouldn't read it."*

We then attempted to present information about their workouts visually. Initially, we introduced a stacked bar chart to represent the daily step count accumulated by steps from each activity, moving away from displaying step data for walking activity alone (see Figure 28 as an example). This approach was also applied to calories (see Figure 30 as an example), minutes (see Figure 31 as an example), and active minutes (see Figure 32 as an example). To enhance their recollection of exercises, we also generated a calendar of their workouts (see Figure 22, Figure 23, Figure 24, and Figure 26 as an example). Additionally, to emphasize participants' daily step count goals, we presented their achieved daily steps in comparison to their set goals (see Figure 29 as an example), particularly addressing feedback from group SP and UP, where the primary aim is to achieve a specific daily step count.

During Visit 3 (see Section 6.1.3.3), we initially presented the visual information described above to participants. Subsequently, we asked them a question about whether the missing day was a regular or irregular day for them (Figure 33) and then revealed the estimated data. Some participants found value in seeing the information used before the estimation, providing insights into the data the algorithm used, as expressed by P10: *"it's good to see everything that was taken into account."* However, others felt that the visual information shown was repetitive, as mentioned by P6: *"I think C6 (active minutes, see Figure 32) is kind of already replicated in C2 (steps, see Figure 28) because you can see that information in C2, but C6 is only specific information on your workout in that two minutes."* Participants expressed concerns about the number of steps involved in this approach, suggesting that they might click next without thoroughly examining the information. P10 stated, *"it's definitely a lot more steps than the other ones. I think that would be*

Figure 27: Explanation of the algorithmic method.



Note: The example above illustrates the text explanation of how the algorithmic method works. It includes a brief introduction to what the algorithmic method is and the source data it references when estimating the missing data.

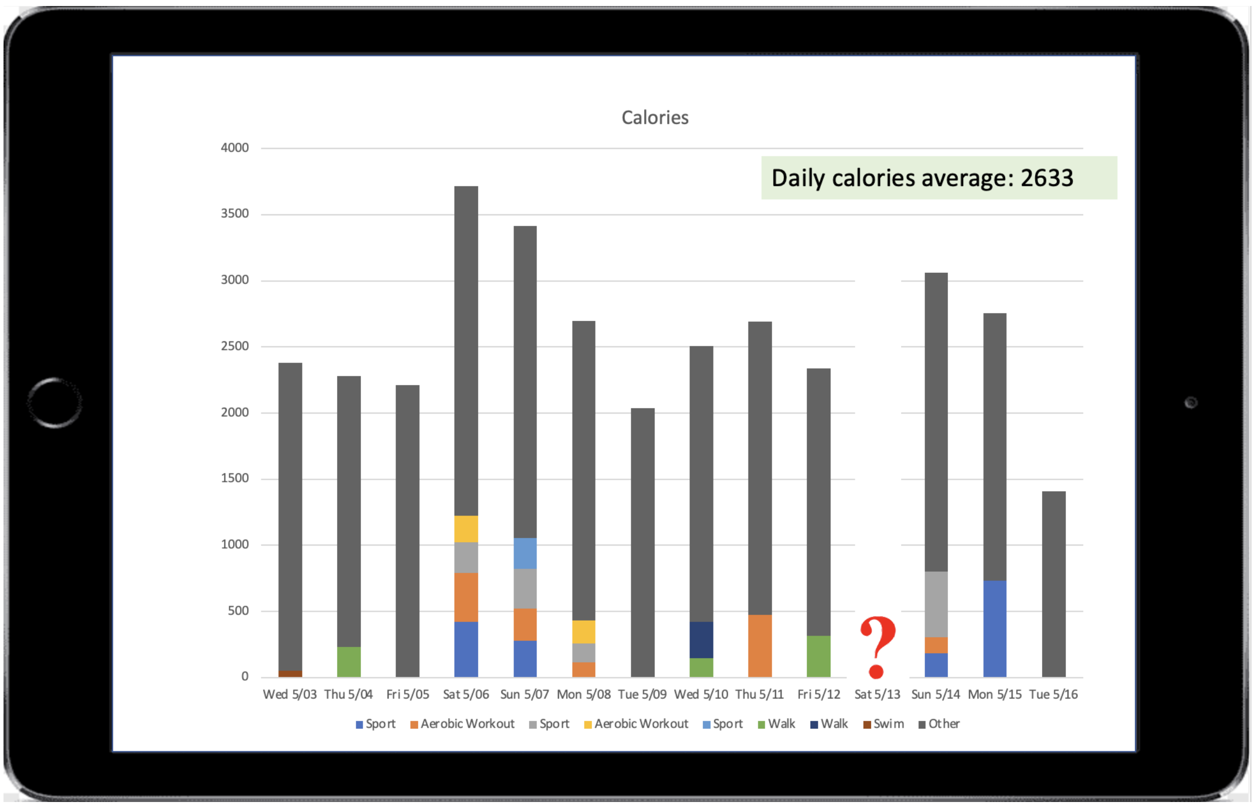
Figure 28: Participants' daily step data.

Note: The above example illustrates the total daily step count, encompassing steps from routine movements, walking, running, and any other physical activities performed throughout the day. For instance, in the step count for 'Mon 5/15,' the blue bar represents steps taken during 'Sport' activities, the grey bar signifies steps from 'Other' activities, where 'Other' denotes steps accumulated during routine daily activities. The question mark stands for the hypothetical missing data, and we also included the daily average step counts.

Figure 29: Participants' daily step data.

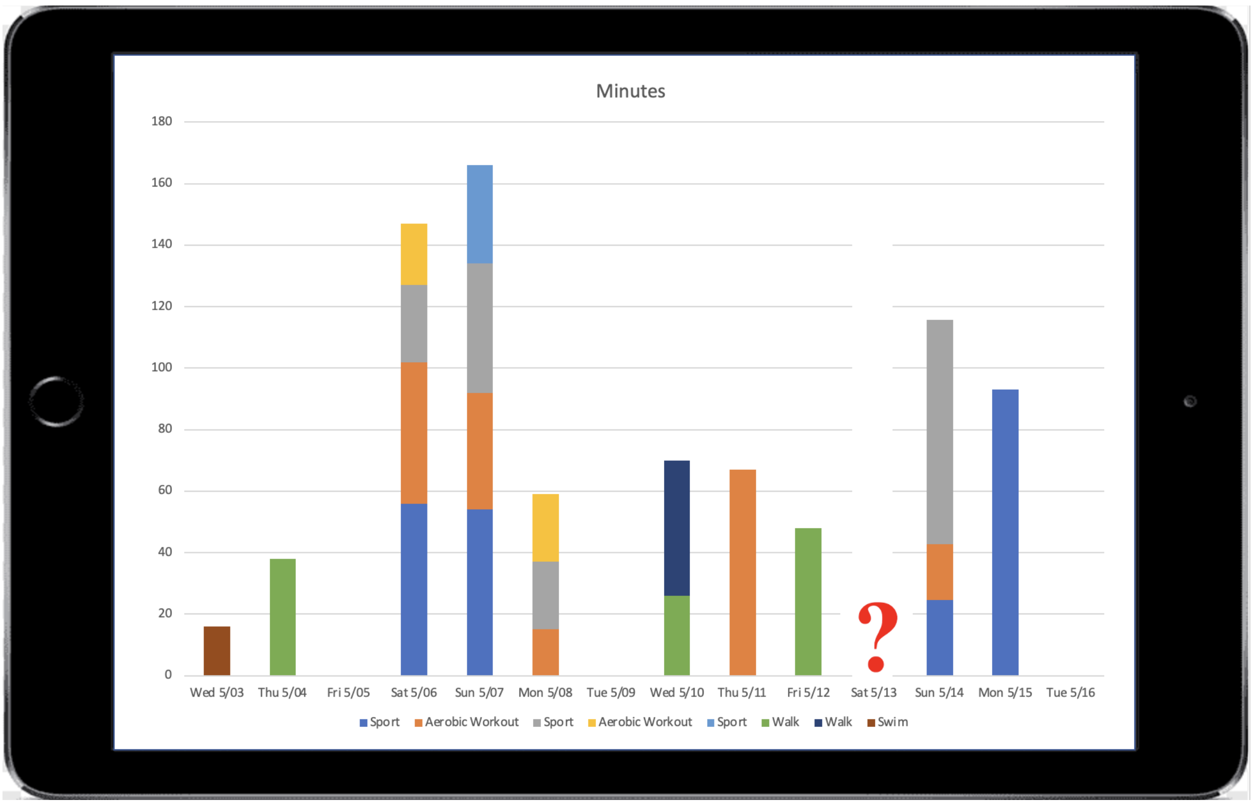
Note: The above example illustrates the total daily step count compared to the step count they aim for. For instance, in the step count for 'Sun 5/21,' the blue bar represents steps taken for the day, while the orange bar signifies steps that are below the daily step count they set. The question mark stands for the hypothetical missing data, and we also included the daily average step counts.

Figure 30: Participants' calorie data.



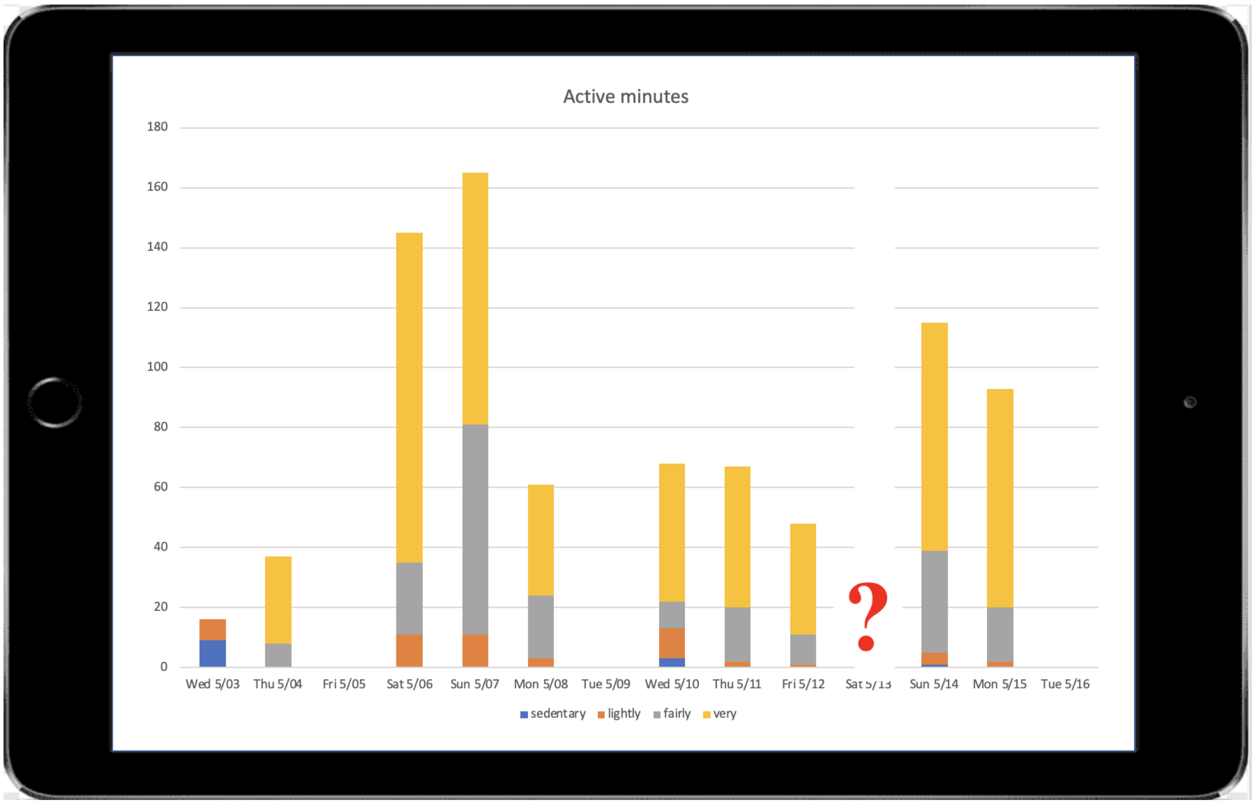
Note: The above example illustrates the total daily calories, encompassing calories from routine movements, walking, running, and any other physical activities performed throughout the day. For instance, in the step count for 'Mon 5/15,' the blue bar represents calories burned during 'Sport' activities, the grey bar signifies calories from 'Other' activities, where 'Other' denotes calories burned during routine daily activities. The question mark stands for the hypothetical missing data, and we also included the daily average calories burned.

Figure 31: Participants' minutes data.



Note: The above example illustrates the time they spent on their workouts, encompassing minutes from walking activities, running, and any other physical activities performed throughout the day. For instance, in the minutes for 'Mon 5/15,' the blue bar represents minutes burned during 'Sport' activities. The question mark stands for the hypothetical missing data.

Figure 32: Participants' active minutes data.



Note: The above example illustrates the active minutes spent on workouts, including minutes from walking activities, running, and any other physical activities performed throughout the day. Active minutes encompass sedentary, lightly, fairly, and very active minutes that are specifically dedicated to intentional exercise. For instance, in the active minutes for 'Mon 5/15', the yellow bar represents "very active" minutes in a day, the grey bar represents "fairly active" minutes throughout the day, the orange bar represents "lightly active" minutes through the day, and the blue bar represents "sedentary" minutes throughout the day. The question mark stands for the hypothetical missing data.

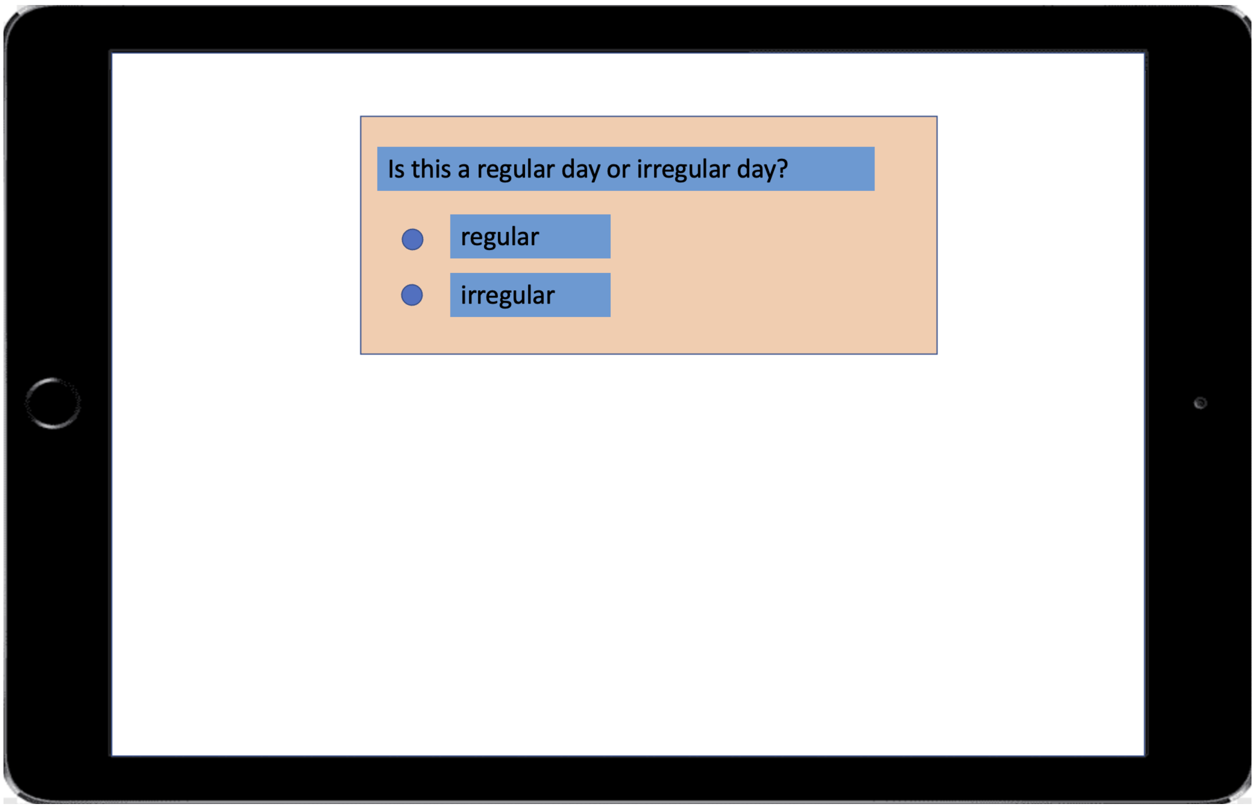
too much to go through that every time." They proposed separating these steps into distinct links for interested users. P10 suggested, *"if you labeled it on one page, like when you go in to get your data, if it just had like bolded, title each of these with a one-sentence description or something, I felt like that would be helpful. So there's like, you know where it (the data) is coming from without needing to see everything."*

Participants expressed appreciation for the prompt (shown in Figure 33), with P2 mentioning, *"it just sort of gives me a little memory boost of what I was doing that day."* P15 stated, *"so that is an irregular day for me, so that has to be taken into account. Otherwise, we will be estimating around 6000 steps, but actually I was in bed. So that is the reason why I like this feature to be added."* Overall, this design increased their confidence in the estimated data, with P6 stating, *"I felt like it worked, what I anticipated that it would look like on the day."* Participants also appreciated seeing their workouts on a calendar. P13 mentioned, *"I really liked the calendar, and how you can click on it and see each thing because that's really breaking down what I did that day. So I kind of know that sometimes if I have a lot of exercise in one day, I can't differentiate what workout was which one. So I like that you could click on it and do it separately."* Additional feedback was provided, such as P13 expressing interest in seeing estimations for calories too.

6.2.5.3 Mixed reaction to event-based approach

During Visit 2 (see Section 6.1.3.2), the event-based approach received positive feedback from most participants, each expressing their liking for various reasons. Among those who liked it was P10, who stated, *"I think it's good. Like if you have data that were similar to it, I know I didn't take this many steps that day, but it was definitely similar to this day here, or somewhere in between."* Others provided positive feedback along with some concerns. P8 mentioned, *"I think it's effective. Except there might be some outliers. Like one day, I might go on vacation."* P12 shared, *"there was a festival going on, and we had an issue with public transportation, so we walked a lot. So it's slightly exaggerated, but more than I would usually do."* P3 noted, *"this was a unique time because I was abroad during this chunk of time."* P2 stated, *"I guess the only thing would be if this day (missing day) had been wildly out of range with my other days, either really low or really high, there's not much option to fill that in."* Some participants preferred the idea of taking an

Figure 33: Prompt inquiry shown to participants.



Note: The figure provided illustrates the prompt inquiring about whether the missing data pertained to a regular or irregular day for the participants. A *regular day* implies that the day was part of their daily routine, while an *irregular day* indicates that something special or out of the ordinary occurred on the day when their data was missing, such as going on holiday.

average of similar days to estimate the missing day, as it didn't heavily rely on their memories. P7 supported this method, saying, *"taking an average of what you did that day and then estimating where it is. That's a very helpful thing. It's better than missing the data completely."* However, not everyone favored the event-based approach, with some participants disliking it due to what they perceived as too much generalization. P12 expressed this sentiment, saying, *"I'm not a big fan of the generalization."*

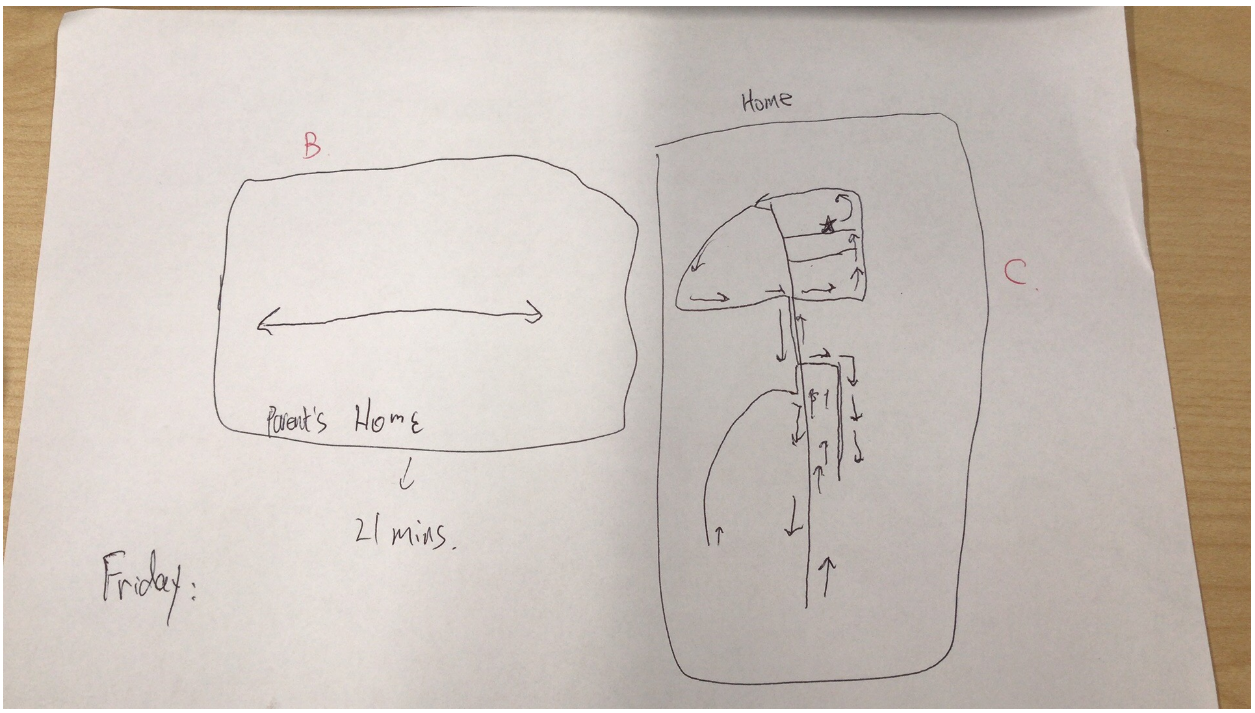
Following the analysis of Visit 2, we observed that participants who prefer using the event-based approach appreciate its lower memory demands compared to the manual-input method. Additionally, they find it valuable to allow them to leverage their own knowledge about how their workouts went, particularly in group SP, where participants have structured routines. To address situations where participants have days significantly different from others, we aimed to provide more flexibility by including options that may not be applicable to everyone (see Figure 36). Furthermore, we incorporated the walking routes participants normally use, which they shared with us during Visit 2 (as depicted in Figure 34 and 35). We also incorporated a feature allowing participants to manually adjust the estimated data.

During Visit 3 (see Section 6.1.3.3), participants' responses to the event-based method exhibited a range of sentiments. Some participants liked it without any criticisms, while others expressed positive feedback with minor suggestions for improvements. Some participants had positive and negative aspects to share, and others disliked the approach and proposed alternative methods.

P4, who expressed satisfaction, stated, *"I feel that it was accurate; there's nothing I don't like about it."* On the other hand, P9 liked it but suggested changes: *"I think it's good that it is taking into account the events because the events are definitely a trigger to the activity level for those days. It might not be absolutely accurate, but it'll be easy to take a guess as to how or what the level of activity was for that."* However, she recommended: *"I also want to see interactive data that could show the most active day, least active day, and what the average activity range is."*

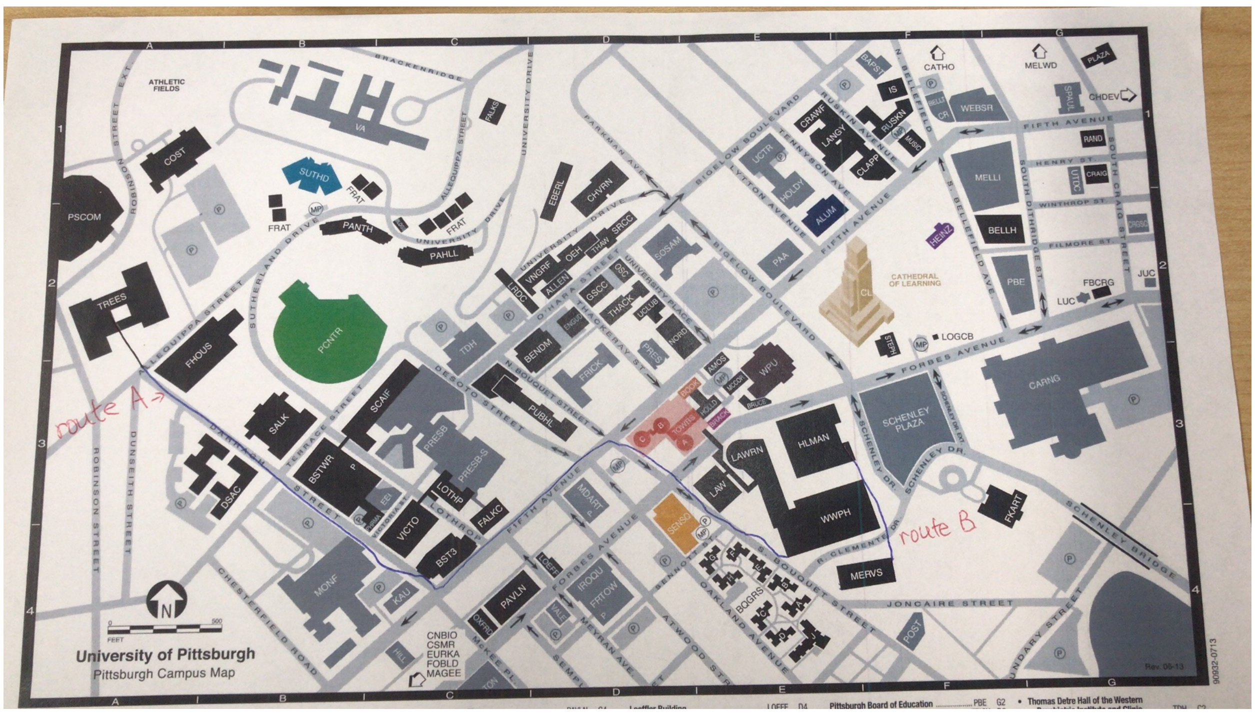
P5 pointed out the similarity in her mind between calories, mileage, and duration, stating, *"if I'm going to have more steps, I'm going to have gone a further distance, burned more calories. So, to me, I would just pick Friday. I'm not going to think too hard about it."* P10 echoed a similar sentiment: *"I walked the same amount these days. They probably burn the same calories and spent the same amount of time walking."*

Figure 34: Participants' exercise route.



Note: The above figure illustrates one of our participant's (a.k.a., P7) route for her walking activities from the iterative user-centered design study.

Figure 35: Participants' exercise route.



Note: The above figure illustrates one of our participants (a.k.a., P9) route for her walking activities from the iterative user-centered design study.

Figure 36: Options shown to participants in MissFit.

Note: The figure above demonstrates the option that we offered as per the request from our participants. This option is provided for cases where none of the presented days are similar with the day when data was missing.

P11 expressed a dislike for recalling activities from a few days ago, saying, *“I don’t pay attention or write about each day or activity on paper or on a calendar.”* P6 emphasized the need for the ability to compare and adjust data, stating, *“if I can compare the day and adjust the data to lower the range, that’ll be good because in my head, I know Monday’s data is a little bit less.”*

P1 expressed uncertainty about the necessity for certain details, stating, *“for the average user, I’m not sure that I need that. But it’s cool to have the option.”* P2 mentioned a dislike for calorie tracking but appreciated having other options, such as maps. P14 felt uncomfortable answering about calories, stating, *“I don’t have a sense about the calorie counts because it’s something that I don’t check.”*

P13 found the process a bit complicated and suggested combining the three methods into one. She proposed, *“maybe this could be an option for that algorithmic versus manual. And then you could have something like this for people who specifically want to get more in-depth about it. But the automatic part just goes without needing to do all these steps every time.”*

6.2.5.4 Collective disapproval of the manual-input approach

After presenting the manual-input approach during Visit 2 (see Section 6.1.3.2 for more details), all participants reported that they disliked it because they couldn’t remember the exact number of steps they had taken. This was especially challenging for participants who didn’t pay much attention to exact step counts. For example, P9 stated, *“I might over or underestimate the number of steps; I don’t remember details about the activity.”* P8 mentioned, *“if I have to remember how many steps I did, that’s beyond me. I have no idea.”* P10 said, *“I don’t remember what times it was between gaps and how many steps, like, that just feels like too much.”* P6 added, *“I don’t really remember how many steps I have.”* P1 expressed, *“that’s a little tricky. I don’t know exactly.”* P14 concluded, *“it’s hard to remember all the steps we take every day.”* Additionally, P14 stated, *“my only concern about this method is my memory could be rough, and the estimate that I’m giving could be completely off.”* In response to inquiries about what would help them refresh their memory on that day, participants responded with suggestions such as reviewing their daily routines, checking calendars, looking at photos taken on that day, recalling special events (i.e., birthdays, friends’ dissertation defenses), etc.

For other feedback on how to improve this approach, participants suggested that instead of asking for specific step counts, the system could inquire about a range of steps. In response to the feedback collected from this visit, we modified our prototype as follows: to provide more information about what their day looks like, we first presented them with a calendar consisting of their workouts (see Figure 22, Figure 23, Figure 24, and Figure 26 as an example). We then presented them with a list of questions reflecting their daily routines (as shown in Figure 37) and asked them to select a range of steps that could represent their missing day or activities (as shown in Figure 38).

We also explored the level of detail that the questions asked, and from the feedback gathered, participants in groups SP and UP predominantly prefer higher-level questions (as shown in Figure 37). In contrast, participants in groups SW and UW tend to favor more detailed questions (as shown in Figure 38).

During Visit 3 (see Section 6.1.3.3), participants appreciated the effort to gather information about their day but suggested improvements, such as personalization, more concise wording, and addressing relevance to individual routines.

P13 expressed a positive outlook, stating, *"I like this because I think that I can trust the data a little bit more."* She believes that providing more information leads to more accurate data but also suggested that the questions should be more personalized, as not all of them are relevant to everyone.

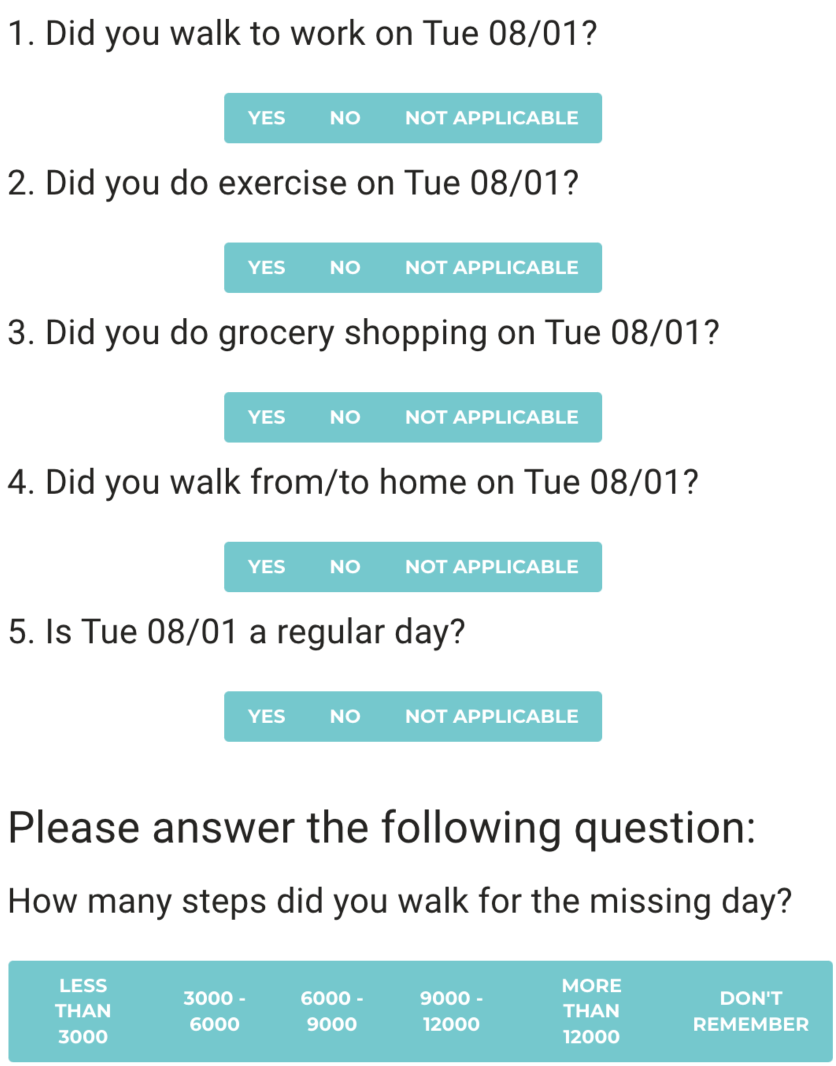
P14 highlighted the memory-intensive nature of answering these questions, saying, *"this is strongly for people who remember what happened during the day."* She recommended adding an option like "I don't know" to accommodate participants who find it challenging to recall details.

P2 observed some irrelevant questions, mentioning, *"I can see that some of the questions are not relevant to me."* She suggested a modification to include questions like "Did you go to more than one location for work or school?" and acknowledged that the questions help in making accurate estimates.

P6 echoed the sentiment that some questions were not applicable to her, mentioning, *"I don't have a lunch break."* She also expressed a dislike for the recall aspect, stating, *"I dislike that it's just a lot of recall. We have to think about it."*

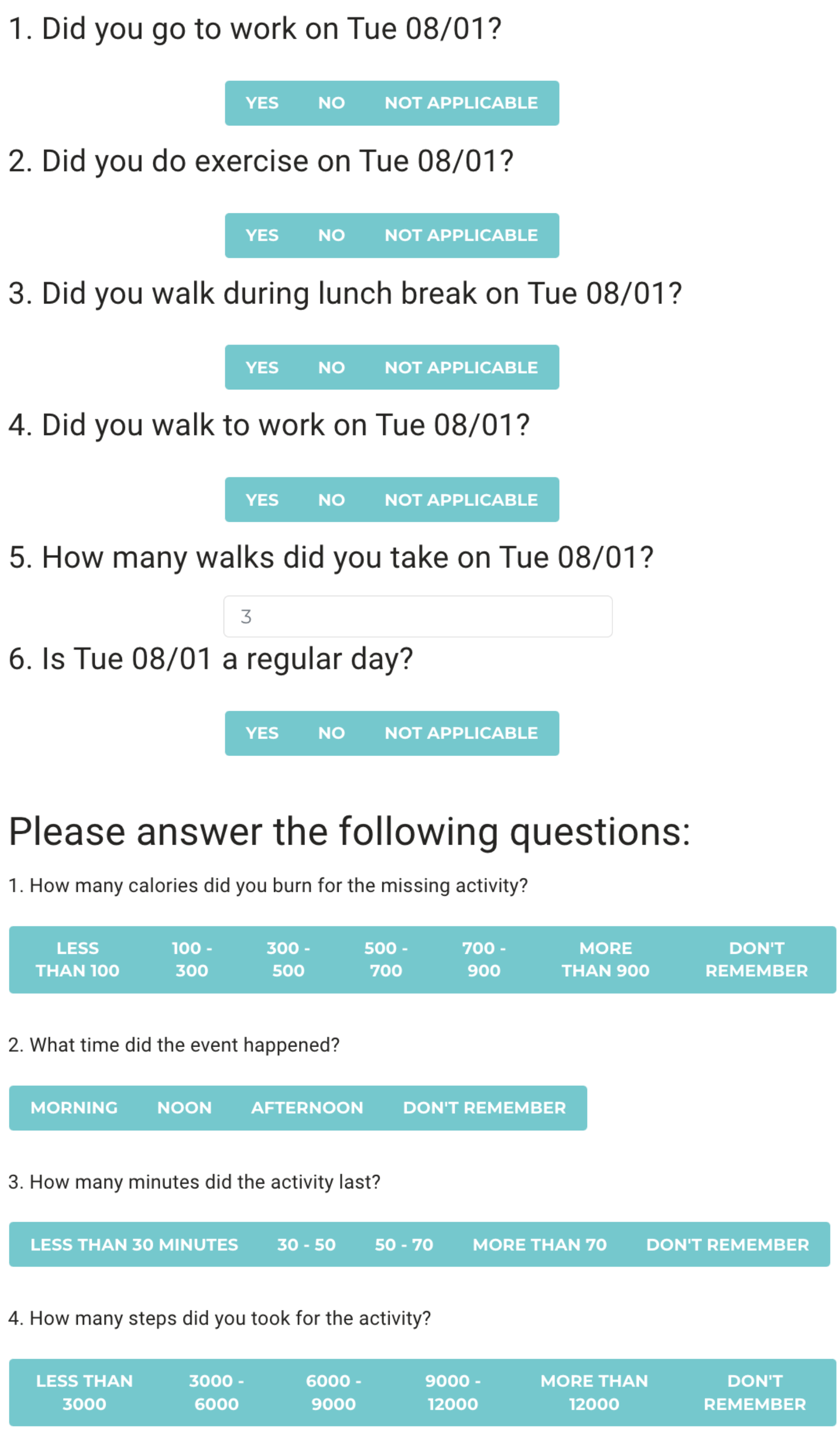
P10 found some questions irrelevant and disliked the recall-intensive nature, stating, *"the lo-*

Figure 37: Questions to help recall the missing day.



Note: The figure above illustrates the questions we asked to assist participants in recalling the day under the scenario where one day's data was missing.

Figure 38: Questions to help recall the missing day.



Note: The figure above illustrates the questions we asked to assist participants in recalling the day under the scenario where one activity's data was missing.

cation option, I didn't go to any of those places. It needs to be either general for everyone or accurate for each person." She expressed concerns about the habit-forming aspect and potential challenges when busy.

P11 cited memory challenges, saying, *"I like the question, but I can't remember what I did, so that's why I cannot answer that accurately."* P12 felt that the questions were not comprehensive enough to cover additional activities beyond a routine day.

On the positive side, P3 appreciated the simplicity of the dichotomous yes-no questions, describing them as easy and quick. P7, however, felt there were too many questions, and the wording made it hard for him to focus. P8 also identified some questions as not relevant to him.

P9 liked the questions that covered most aspects of her day, stating, *"I think he covered most of the things that I would have done."* She appreciated the broad coverage, especially regarding activities after school.

There's also a disagreement on the appropriate number of questions presented. P13 expressed a preference for a specific range, stating, *"from 3 to 11 is the acceptable range for the number of questions."* She emphasized that exceeding 10 or 11 questions could be perceived as excessive. On the other hand, P2 had a broader acceptable range, mentioning, *"less than 20, it didn't bother me."* She also pointed out that too few questions might not provide sufficient information, she likes to utilize her own knowledge, and she likes to answer questions in general. P3 found seven questions to be suitable, expressing that this number felt appropriate for the method. These diverse perspectives highlight the need for personalization and further research to determine an optimal number of questions that balance obtaining valuable information without overwhelming participants.

6.2.5.5 Design evolution through visits

Due to the nature of the iterative study, participants were recruited at different times. In each iteration, we incorporated global feedback from all participants. For example, we recruited 4 participants in week 1 and 3 participants in week 2. The design feedback received from the first 4 participants influenced the design for the 3 participants in week 2. The design gradually evolved as we collected feedback from all our participants. This process contributed to the observation that most of our design changes occurred during Visit 2, and the design started to converge during Visit

4.

During visit 2, we identified that participants were interested in understanding how the algorithmic approach generated the estimates, that the event-based method needs to be less generative, and that the manual-input method needs to be less memory intensive. During Visit 3, we explored different options to mediate the problems identified during Visit 2. From the feedback, we identified participants favor systems that are easy to use most of the time but also want the freedom to check details and utilize their personal knowledge. Indicting a converged design that combines the three methods into one. During visit 4, we tested the converged system and new requirements from participants. This converge happened at the end of our recruitment, which means fewer participants provided feedback on this version.

7.0 Implications

7.1 Design implication 1: systems should utilize the users' daily habits to guide them to recall the missing data

From the iterative user-centered design study (outlined in Section 6.2), we learned that remembering specific numerical values or activities is challenging for participants, but recalling using their daily routines is more manageable. This is particularly evident for participants in groups UW (described in Section 6.2.3) and UP (described in Section 6.2.4). For example, P8, a member of group UP, doesn't have a predefined exercise routine, *"I usually walk to go to my classes. And if I'm not at a decent amount of steps, I'll walk a little more to reach those steps. So it (the routine) depends on the day."* When the manual input methods required her to recall step counts, she preferred alternative designs, stating, *"I just have the worst memory in the world."* She suggested that methods assisting her in mentally walking through her day would be beneficial, *"it's easier for me to remember my day when I can walk through it. If it's like asking me to recall specific parts of my day, that kind of extra ahead. But if I could walk through the day, in my mind, you'd have questions that would facilitate that. "*.

P11 emphasized the difficulty of recall, stating, *"The biggest problem with anyone is the power to recall."* Three approaches utilizing daily routines were identified to aid participants in recalling the missing day. The first involves presenting users with a calendar displaying all their activities. P8 found the calendar helpful in mentally walking through her day, saying, *"This (the calendar) helps me walk through it as well, so this is a good representation of how your activities might be."* P13 expressed a positive view of the calendar, appreciating the breakdown of activities, saying, *"I liked the calendar, and how you can click on it and see each thing because it breaks down what I did that day."* P10 found the calendar's information aligned with what she typically examined. P5 said, *"I can refer based on what it showed. I played for a long part of the day, then that was probably a tryout, and then based on that, I could guess what the next day was."* and P11 said, *" I think it's easier to like looking at the calendar, then you can help yourself by eliminating certain period, like 10 am to 6 pm is when I work, which means that time slots was gone, and then you can* *calculate the other things."*

The second approach involves asking questions about participants' daily habits. P8 suggested questions like, *"Did you go to work? Did you go to school? Did you go to soccer practice? Did you go to the park?"* P7 appreciates those questions, commented, *"They are not just questions, I guess it's more about the thinking when you answer the questions."* P9 appreciated the specificity of these questions presented during visit 2 (see details in Section 6.2.5.4), stating, *"I think it (the questions) is specific, and it's breaking it down into different levels of your day, like getting into school, getting a lunch break. Things you do after school are easier because it begins to zoom in and see the different activities for a particular day."*. And P10 recommended asking simple questions about daily routines, *"ask simple things, like yes, they went to work, then (the system should) show them when they're at work this is how many steps they tend to take in a day."*

The questions should be personalized so that they are relevant to each user. P6, while she did not find the lunch break question personally relevant to her, she acknowledged its importance for those with a routine lunch break, *"I think if somebody routinely does a lunch break, that would be an important question, but for me specifically, it's not as important."* P14 suggested reducing the number of questions and including options for cases when participants couldn't remember certain details, *"instead 'Yes' or 'No' options, it should also add 'N/A' option in cases when the question doesn't apply to them or if they don't remember."* P13 favored the questions, as they provided more information, increasing trust in the synthetic data generated, *"Because I'm giving you a lot of information. So I trust the synthetic data that comes out is pretty spot on compared to not having any of that."* She also suggested, *"The questions are good. But I think some of the questions are irrelevant to some people, so it's better only to ask relevant questions."*

The third approach involves utilizing map locations, considering that participants frequently visit the same places or use familiar routes for walking, running, or walking activities. So, the map locations can be a good way of helping people recall their activities. P1 noted that occasional variations in routines, such as going to a different park, could make some days challenging to generalize, *"On Saturday, we just happened to go to a different park that we don't usually go because we went for a child's birthday party. So yeah, some days are just a little different."* P6 suggested the option to select multiple locations for days spent in other neighborhoods, emphasizing the relevance of this feature for diverse scenarios, *"If I was at two different locations that day, it might* *be important to be able to select both."*. P10 preferred the map location method, indicating that it would be more suitable for her.

P12 highlighted the consistency of walking activities in her neighborhood. At the same time, P7 identified two primary locations, *"I am normally at two locations: one at home and one when I walk the dogs. I do the same walk in the morning and at lunchtime."* Overall, these three approaches leveraging calendars, targeted questions and map locations offer promising avenues for assisting participants in recalling missing data based on their daily routines. Future PI tools aiming to mediate missing data should incorporate these features to help end users recall their missing data.

7.2 Design implication 2: systems should utilize the users' daily routines in estimating their data

In the iterative user-centered design study (refer to Section 6 for details), we discovered that participants' exercise routines are linked to their daily schedules, which vary among participants. The estimation methods should consider incorporating participants' daily routines to enhance the accuracy of estimating missing data. This connection is particularly evident in the behaviors of participants in groups SW (see Section 6.2.1 for details) and SP (see Section 6.2.2 for details). For instance, P7 from group SW described a well-defined exercise routine integrated with daily activities: *"I'm in the office three days a week. So every morning, I try to get here early and walk for about 20 to 30 minutes. And then, at lunchtime, if I can, I try to walk for even an additional 20 to 30 minutes. And then pretty much every day after work, around four to five, I walk for 20 minutes, playing outside with my dogs, walking back and forth to get the ball to them."* Similarly, P12 in the same group reported, *"I do them every morning. The first thing I do when I wake up, I do about 45 minutes of cardio, and then I do resistance training like lifting. I do another 45. So I'm at the gym for a good hour and a half on a weekday. On weekends, I'm a little bit longer, usually, but roughly an hour and a half between the cardio, like half cardio half resistance training."* However, each participant's routines are different. For example, P6 in group SP mentioned, *"I did go to the beach again for a couple of days. So my walking location would be a little bit different. Other* *than that, there were only what? Four days? I was there for the holidays. I was there for the show. So those days are just different locations."* P14, also in group SP, added, *"In summer, I make it intentional, I try to walk back home where I would take some 4000 steps extra. So during the winter, it is a bit hard for me to maintain that, but a minimum of 5000 is what I try to maintain. And only from Monday to Friday."* Furthermore, routines fluctuate during special events.

We identified two ways to leverage participants' daily routines to aid the estimation process. The first approach involves integrating user input into the algorithmic method by prompting participants to specify whether the missing data corresponds to a regular or irregular day. Regular days align with participants' daily routines, whereas irregular days indicate deviations or special events. P13 articulated a preference for this feature, stating, *"I would want the last option that we had, asking if this was a regular day or an irregular day. And then if I clicked a regular day, then I could manually put more information in to make it more specific. But if it was a irregular day, or I know I did a lot of exercise that day, I could say no, it was an irregular day and then manually put more information in."* P14 supported this perspective, highlighting the importance of accounting for irregularities, such as days with minimal walking due to specific circumstances, *"like what happened to me on a Tuesday, I didn't walk much, hardly 1000 steps. So that is an irregular day for me, so that has to be taken into account. Otherwise, we will be estimating around 6000 steps, but actually I was in bed. So that is the reason why I like this feature to be added."*. P2 expressed approval, recalling a positive experience with a similar algorithm that distinguished regular and irregular days based on the participant's activity level. "I like it. For example, I know when thing was out of the ordinary because I helped my friend move so my heart rate was really up that day. So I would know that it was irregular for me." P6 identified a vacation day as an example of an irregular day, emphasizing increased physical exercise. *"I would say this is probably an irregular day because I was on vacation. So I was doing more physical exercise and walking I normally would do."* P10 reflected on the recent past, noting changes in routine during travel and adverse weather conditions. *"I feel like this has been more of my regular routine than going back into last weekend. I walked a lot more because I was traveling and seeing a different city. And then also the weather too. It Was raining a lot. I've definitely been outside less."* P4 acknowledged the regularity of exercise routines during weekends compared to weekdays, *"my exercises routines are more regular during weekends compared with weekdays."*

The second approach involves ensuring that the event-based method options capture various aspects - personalization for each user. For example, P1 expressed satisfaction with the event-based method, stating, *"I appreciated the event-based method; I have a good sense of what my day looks like, and I don't think I need that level of detail (manual input) to track my calories accurately."* P10 presented a different viewpoint, saying, *"Yes or no questions are easier than choosing similar days. I walked the same amount on these days, so they probably burned the same calories, and I spent the same amount of time walking. The questions should capture different aspects."* P13 proposed an alternative, suggesting, *"by clicking on the bar itself to select specific days is easier."* P14 provided a balanced perspective, stating, *"I think it's good if I can recall. But we should consider adding one more option like 'that day is not similar to any of them.'"* P9 expressed a desire for additional features, stating, *"I would like to see data and interactive information that shows the most active day, the least active day, and the average activity range."* P2 added, *"I don't pay as much attention to the calories as to other stats. I didn't dislike the calories. As I mentioned, some people notice different things. I actually think I would like it if it addressed both. I generally prefer more options. The more data it asks for, the better I would feel."*

Additionally, both approaches ensure that the methods reflect changes in participants' daily routines- personalization- and the algorithmic models that generate the estimated data should be adaptive. For instance, P6 highlighted the challenge of using routine schedules during an atypical week, suggesting that routine-based comparisons would be more effective if activities remained consistent week-to-week. *"If I were following a routine schedule and engaging in the same activities last week as I did this week, using data from last week would be beneficial. However, since my activities were different this week, comparing my current data with last week wasn't helpful. I found it challenging to select a specific day because some days lacked any information or walking data. Consequently, I chose days with the highest values, as I was more physically active during my beach visit. I believe in a consistent routine, utilizing last Monday's information for this Monday could be valuable. So, I don't think it's something to discard. I think it was an unusual week for me to use that feature effectively."*

Participants' exercise routines are intricately linked to their daily schedules, emphasizing the need for estimation methods to consider them. The two identified approaches—incorporating user input into the algorithmic method, refining options for the event-based method, and ensuring adaptability to changes in routines—provide promising strategies for enhancing the accuracy of missing data estimation. Future developments in PI tools should explore integrating these features to better assist end-users in estimating their missing data.

7.3 Design implication 3: systems should prioritize cumulative measurements but still offer details of individual discrete activities

The system should prioritize cumulative measurements while offering detailed information on individual discrete activities. From both the semi-structured user study (see Section 4.2) and the iterative user-centered design study (see Section 6.2), we found that participants, referred to as *maintainers*, who have a measurement-focused personality, strongly prefer having an overall measurement of their daily activities. Participants with a wellness-oriented personality prefer to receive a general measurement of their daily activity level and a tendency to check specific details of their discrete activities.

In the study, all participants, regardless of their personality types, utilize one or multiple cumulative measurements to measure their daily activity level. Specifically, participants fitting the persona of Group SP (see Section 6.2.2) and Group UP (see Section 6.2.4) aim to maintain a relatively active lifestyle and strive to meet or exceed specific step counts or exercise durations. Their primary focus is confirming whether they have achieved their daily activity targets, with less emphasis on examining the finer details of their discrete activities. However, the matrix that each participant uses is different, which calls for personalization. For example, P6 from group SP said, *"I just look at my steps. The 10,000-step goal is usually what I look at."* It's worth noting that participants employ various cumulative measurements to represent their daily activity levels. P14, from group SP, said, *"my goal for using Fitbit is to see how many more minutes I have and how many steps I take in a day. So know whether I'm active enough for a day and not sitting in one place."* For instance, P1 uses Fitbit to reach a daily step target of 10,000. However, she observed that weightlifting was not counted toward this goal. So, she had to switch to other exercises, leading to inconvenience.

For those fitting the Group SW (see Section6.2.1) and Group UW (see Section6.2.3) persona, they track their progress toward specific performance goals. They exhibit greater attention to individual discrete activities, although they acknowledge that probing into these details every time may require too much effort. For example, P9, from group SW, said, *"I mainly use the daily steps and the calories to track my performance."* And P2, from group UW, said *"I think links for more detailed information is helpful, like overall information, but also details about each activity, maybe not every time to go through it, but I would like that option."* Throughout our study, we noted that certain types of exercises, such as weightlifting and cycling, were not factored into the overall health measurement, causing burdens to participants. P7 commented, *"I pay more attention to metrics like active minutes and heart rate since they better reflect my daily performance. For example, in the summer, I engage in intense upper-body workouts that don't register many steps but are good workouts for me."* Similarly, P13 highlighted the current system's limitations, stating, *"If I'm in stationary, like lifting without much movement, it doesn't register steps. However, dynamic exercises like lunges or squats involving leg movement are counted. Standing in one position for lifting doesn't register steps because the system needs leg movement to detect steps."*

One approach to this is the breakdown illustrating how various activities contribute to the overall measurement. In our study, participants were presented with a stacked bar chart depicting daily step counts (refer to Figure 28), calories (refer to Figure 30), active minutes (see Figure 32), and minutes (see Figure 31), eliciting positive feedback. For instance, P12 expressed, *"I like seeing the breakdown for steps, calories, and minutes."* P3 commented, *"I prefer them stacked rather than just side by side."* Additionally, P11 remarked, *"I prefer this one (breakdown for each walking activity). I think this one is more accurate; it shows the steps I took from each activity."* P13 stated, *"I really liked the calendar, and how you can click on it and see each thing because it's really breaking down what I did that day."*

Another approach involves presenting a summary list of their daily activities, with the option to dig into specific details by clicking on the activity. This was tested by condensing daily activities into a calendar format. P5 found this feature helpful, stating, *"this (the calendar) would be helpful to me."* Similarly, P11 expressed, *"I don't pay attention or don't write about each day or activity on a calendar, so I like it that when I click on the green square (the activity), it gives a breakdown of heartbeat, energy burn, etc."*

In addition, providing explanations on how each measurement is calculated can enhance participants' trust. For instance, P10 noted, *"I'm somebody who doesn't pay attention to calories. Like when I cycle, it says I burn like 300 calories, but I have no idea how it (the calories) was calculated."* Similarly, P14 expressed, *"As I told you, I have less confidence (about calories)."* P2 mentioned, *"I don't pay as much attention to the calories as to other stats. I have little idea how calories worked."* Additionally, P8 remarked, *"I feel you can't really apply calories as much as minutes."* Some participants held suspicions about the data collected by their wearable devices, leading them to withhold from using the measurement matrix itself (illustrated in Section 6.2.5.3). For instance, one participant mentioned not using calorie data because she didn't understand its value, stating, *"when I woke up in the morning, it (the Fitbit) shows that my calories are already 50 something. I mean, where does the number come from?"* Another participant questioned the difference between daily step count and walking steps. Some participants also noted discrepancies in their objective to exceed a specific daily step count, as they observed that some exercises were not counted toward the daily steps. All of these instances hindered trust in the system itself.

7.4 Design implication 4: systems should utilize prompts to identify and estimate missing activities

Users' feedback can be utilized in three ways: the first is to identify missing activities to initiate the estimation process, the second is to help build estimation models, and the third is to help guide the estimation process, and personalization is apparent for all of them. From the analysis of the algorithmic approach (discussed in Section 3.6), we learned that it's challenging for algorithms to identify missing activities accurately without human intervention. For example, if the missing activity occurred at night, the algorithm could identify it as morning. This could undermine participants' trust in the system. However, identifying missing data and days is the initial step in estimating them. During the semi-structured study (outlined in Section 4), we identified that maintainers displayed a consistent pattern of utilizing and actively engaging with the tools at the time of data collection, often to confirm data acquisition and monitor progress toward their goals (see Section 4.2.2).

So, the system could prompt users with confirmation, like 'is this a missing activity/day?', with

'Yes' or 'No' options. It can be combined with the auto-detection of the activities to identify missing or inaccurate activities or days. Current devices can detect whether participants are engaging in a specific activity, but it may result in undetected minutes or errors if participants work out at the gym. For instance, one participant in the semi-structured user study reported that when she uses auto-detection on her watch, the watch typically detects the activity a few minutes late, leading to inaccuracies in the record. Other participants also reported that sometimes auto-detection provides an incorrect activity type. This might be a suitable time to prompt users with a question to confirm the collected data. Another appropriate time would be during their reflection on their goal-tracking process, such as a daily reflection on data collected, checking for missing activities, or instances where data wasn't recorded. This approach can improve participants' confidence in the collected data and initiate the estimation process. It can also have prompts that ask users to identify activity patterns at these times, such as 'over the last month or week, is this a regular/irregular day?'. Later, the system can utilize the data from the regular or irregular days to build the estimation model.

During the estimation process, the system can prompt users to determine whether the missing day is regular or irregular. If it's a regular day, the system can refer to the model constructed from regular days to provide an estimation. However, if it's an irregular day, the system should further prompt the user with questions such as 'how long is the activity?' 'what's your estimated range of the missing data?' 'does the missing day have comparable days?' etc. Afterward, the algorithm would generate an estimate based on the user's input to these questions. As expressed by P13, *"if I clicked an irregular day, then I could manually put more information in to make it more specific. But if it was a regular day, I can trust that this was average from my other regular days."* P14 appreciated this approach, stating, *"I like it that it's asking that question whether it was regular or irregular. I like this approach, giving the user the opportunity to get involved in the algorithmic process rather than just presenting the result."* Similarly, P2 commented, *"I like how it asks if it's a regular or irregular day because I do have days that were out of the ordinary, then I would know that it was irregular for me."*

After presenting the estimated data, the system can prompt users to determine if they want to make further adjustments. During the iterative user-centered design study, we found that participants often have their expectations regarding the estimated data. For example, P10 said, *"I know I didn't take this many steps that day, but it was definitely similar to this day here or somewhere* *in between."* and P2 said, *"I guess the only thing would be if this day (missing day) had been wildly out of range with my other days, either really low or really high, there are not many options to fill that in."* We also found that participants hold diverse expectations regarding the estimated data. Some participants anticipate the estimated data will be close to their daily averages. Some participants expect the estimated data to follow similar patterns as in previous days. For instance, if the missing data corresponds to a regular day during the weekend, they would expect the estimated data to exhibit similarities with previous weekend days. By ensuring that the estimated data is consistent with participants' expectations, we can enhance their trust in the accuracy and reliability of the estimations.

We explored two different approaches. The first approach can be interactive, providing a slider for the estimated value, and participants can slide up or down to increase or decrease the estimated data. P2 expressed, *"I like it."* P14 stated, *"I like that you are giving the user the option to change in case they remember what the missing data is, and they can say no if they just believe what the algorithm gives them."* The second approach can be facilitated by the event-based approach, presenting users with comparable days to the missing day and allowing them to adjust the data on a list of comparable days. P6 mentioned, *"if I wanted to compare my Monday (missing day) to Tuesday, and if I can lower the range a little bit somehow because I know my Monday is less than Tuesday, that would be helpful."* P3 commented, *"I like the option that I can choose to adjust the estimated data, but it also made me wonder if there is a reason for me to want to adjust it. But I like the option."*

7.5 Design implication 5: Systems should help users understand the estimation methods used by providing a complete and detailed summary of the "source" activities referenced

Systems should help users understand the estimation methods by providing a complete and detailed summary of activities referenced to generate the estimated data if the estimation method is not intuitive. We learned from the iterative user-centered design study (see Section 6.2.5.2) that participants found the event-based approach, which utilizes events and comparable days to generate an average for missing data, and the manual-input method, which relies on users' memory to fill in missing data, were perceived as intuitive and garnered different preferences. However, the algorithmic approach raised concerns. For instance, participants mentioned that they lacked trust in the estimated data because the process did not provide them with the information needed to comprehend how the algorithmic approach operates. For example, P1 mentioned, *"That seems odd to me because why would the algorithm assume that I had more exercise than the day that was missing?"* P9 said, *"I like that it (the algorithm) is not requiring me to go down memory lane to try and remember, but I also don't know if what it is doing is correct or not."* P3 said, *"I'm not sure how they came up with that number."* P4 questioned, *"I think the estimated data is wrong. How did it generate this?"* And P7 said, *"the algorithmic approach is definitely more accurate, but I don't know, the algorithm is boring. How is it generating this (the estimated data)?"*

First, it is crucial to inform participants about what has already been recorded to estimate missing or partially missing data. During the study, we presented visual information such as calories burned, specific activities they engaged in, and active minutes before introducing the estimation methods. Participants believed that the estimation methods used the presented information to guide the generation and reported that this information helped them build more trust in the estimated data. For example, P3 stated, *"Having more details about my activities helps me understand and build more trust in it (the system)."* P8 said, *"Because I have my calories here as evidence, I trust it."* However, this visual information should be personalized based on participants' interests because participants focus on different aspects. For example, P8 said, *"The calories presented here gave me more confidence, but if it were to offer me minutes, I would say no because I'm more aware of my calories."* P9 said, *"During active minutes, you are going to burn more calories; you can see how it directly correlates with active minutes, so the active minutes seem like extra information, and I only care about calories, so just showing calories is enough for me."* By presenting users with the known data, they can make informed estimations and fill in the gaps based on the available information. This transparency and visibility into the recorded data empower users to make more accurate assessments, enhancing their confidence in the estimation process. P10 said, *"It's good to see everything that was taken into account."*

Secondly, the system should explain the methodology employed in generating the estimated data, providing either the data used or a descriptive overview of how the method functions. This description should be high-level, supplemented with additional links that users can click to inspect the data used to generate the estimated data. This approach addresses the tendency of users to bypass detailed text explanations of the method's functionality. As reported by P5, *"Even if it told me how it generates the data, I'd probably just hit 'next'."* In agreement, P3 expressed, *"Summaries like 'the algorithm uses your historical data to help guide the generation of the estimated data' are sufficient."* Conversely, some users suggested a more detailed explanation, containing specifics like *"what types of data the algorithm uses and what dates it considers to help generate the estimated data"* would be beneficial. P10 offered a suggestion, stating, *"It's good to see everything that was taken into account, but I think that would be too much to go through every time. It's helpful to put them in separate links so people can check them when they want to."*

7.6 Design implication 6: systems should trade-off between comprehensiveness and ease of use when estimating the missing data

From the iterative user-centered design study, we found that participants valued comprehensiveness in the estimation methods. For instance, they appreciated including more questions in the manual input method to help them recall the missing data. P11 believed more questions prompted better recall and choices, saying, *"I think if somebody has more questions, they will try to recall their activities for that day better. Like if there are more questions and with more options, then I'll try to recall what I did myself so that I can choose options."* Similarly, P2 said, *"I feel if there are too few questions, I don't know that I would provide enough information to it (the estimation method)."*

The study also found that explaining the details thoroughly can increase participants' trust in the estimated data. For example, in the algorithmic approach, participants reported increased trust in the result when we listed the historical information used to generate the estimated data. P10 stated, *"Seeing where it's (estimated data) coming from makes me trust it a little bit more."* In the event-based method, P2 also claimed, *"The number of questions makes me feel I trust it more than if there had been fewer (questions)."* P13 said, *"Because you are asking more information about my day, the more knowledge I give you guys means the more accurate the data will be, so I trust it more."*

However, we also found that participants favor methods that require minimal effort simultaneously. For example, some participants don't want to answer all these questions in the manual input method. P5 said, *"I don't want to have to put in these questions because, as a user of apps, I would want it to just calculate it for me. And I would just trust that it's doing it."* Participants felt the event-based method contained too many steps when providing detailed information about the historical data. For example, P13 said, *"For people who don't care as much, they would probably just want it to be no work. Adding an extra step could turn away some people who don't care as much. But for people who do care about it and use the Fitbit for specific purposes, they wouldn't mind the extra step."* In the algorithmic approach, participants also felt there were too many steps added. They didn't want to go through all the steps after a long day. For example, P7 said, *"There are too many steps. I would try to do it once or twice maybe, but not frequently, probably not after a busy day at work."*

We identified a process to combine three methods, allowing users to decide which step to stop. The system will apply the algorithmic method by default upon starting the estimation process. After the user identifies whether the missing data is from a 'Regular' or 'Irregular' day, the algorithmic method presents the estimated data with links explaining the data points the method referenced. The system will then ask whether users want to adjust the estimated data. Users can use the event-based or manual-input method to provide more information about the missing data and refine the estimated data. If users are satisfied with the algorithmic methods or don't want to go through extra steps, they can stop with the estimated data generated by default.

During the study, some participants liked the process for its capabilities. P14 said, *"The different options given make me feel that they can help the system with better accurate prediction."* P13 claims, *"I really like that. So if you get to say it's a regular day, and then just stop there. Or you can say it's an irregular day and go through the event or manual, based on which one fits you. I like having the option."*

Some participants liked it for more control and flexibility over the system. For example, P14 said, *"I feel this is good. And it gives me more freedom to use my memory or go with this. So people who don't remember this can just go with the default options; those who remember, let them iterate and figure out what they remember."* P2 said, *"I like it that you have options to just things either way because maybe sometimes you remember more specifics, and other times you remember, not* *at all specifics, but you're like 'I think it was similar to one day.' So that's nice because it gives you more flexibility."*

When asked if participants wanted to adjust the value, some said no. P13 claims, *"I said 'No' because I know it's a regular day. But if I had set goals I was working for, I would want to adjust it and make sure it's correct, so I like that feature."* P13 claims, *"I wouldn't want to go through that process every single time. So I would like to have the option of saying 'No,' but I like that if I wanted more than that, I could pick 'Yes' and choose to put more information in."*

7.7 Discussion

7.7.1 Connection of our design implications to other fields

This thesis addresses a crucial aspect of Personal Informatics (PI) systems—the issue of missing data and how users perceive and resolve it. The focus on understanding user perspectives on missing data and exploring methods to address these concerns is essential for designing more effective and user-friendly PI systems. These studies contribute valuable insights into user needs and preferences, which are critical for guiding the development of future systems. By proposing design implications, the thesis provides a foundational framework that can help designers create PI systems that more effectively mediate and explain missing data. The observation that user trust in the system significantly depends on their understanding of how estimation methods generate data highlights the importance of explainable systems. This connection to Explainable AI (XAI) is particularly pertinent.

XAI refers to artificial intelligence systems designed to be transparent, providing human-understandable explanations for how they arrived at their decisions or predictions [3, 33]. The goal of XAI is to clarify AI's decision-making processes, enabling users to understand, trust, and effectively manage AI technologies [3]. This is particularly crucial in sensitive or impactful areas like healthcare, finance, and autonomous vehicles, where understanding the rationale behind an AI's decision can be critical for acceptance, ethical considerations, and regulatory compliance [3, 8, 35]. This area has led to the development of design implications that offer high-level guidance on how to build more symbiotic systems between reasoning systems and humans, specifically the following:

- **User-centered design**: design explanations with the end-user in mind, considering their background, expertise, and what they need to know to make informed decisions or understand AI outputs. Tailor explanations to the level of technical expertise of the target audience [34].
- **Transparency**: Ensure that the AI system provides clear and accessible information about how decisions are made. This includes details about the data used, the decision-making process, and the logic behind specific outcomes [66].
- **Simplicity and Understandability**: Use simple and intuitive explanations. Avoid technical jargon and present information in a way that is easy to understand. Visualizations, such as graphs, decision trees, or heat maps, can help users grasp complex concepts more easily [69].
- **Contextualization**: Provide context for AI decisions, including why a particular output was generated and how it relates to the input data. This helps users understand the relevance and applicability of AI decisions in real-world contexts [82].
- **Feedback Mechanism**: Incorporate mechanisms for users to provide feedback on the quality of explanations and the decisions made by the AI system. This feedback can be used to continually improve the explainability and overall performance of the system [58].
- **Multimodal Explanations**: Consider offering explanations in multiple formats (text, visualizations, examples, etc.) to accommodate different learning styles and preferences. This can enhance the overall accessibility and effectiveness of the explanations [1].

A significant portion of the design implications is well represented in our work. Personalization is a crucial aspect of our design implications, aligning them with the principles of user-centered design. Future Personal Informatics (PI) tools should be designed considering users' backgrounds, expertise, and needs. Design implication 5 (see Section 7.5) is aligned with principles such as transparency, simplicity and understandability, contextualization, and multimodal explanations. This suggests that future PI tools should help users understand how the estimation method works, with explanations provided in multiple formats that are easy to comprehend. Similarly, design implication 4 (see Section 7.4) aligns with the principle of incorporating a feedback mechanism. This implies that future PI tools can utilize prompts as a feedback mechanism to identify and estimate missing activities.

7.7.2 Connection of our work to Personal Informatics

In this thesis, we present an initial exploration into how users interact with and perceive missing data within the Personal Informatics (PI) system. Focusing on user-centered design, these studies delve into user perceptions, the impact of missing data on user experience, the natural behaviors users have when they encounter missing data, and design implications future PI tools should incorporate when designing systems that mediate missing data. Our work contributes to the PI field in the following ways:

- **Understanding of user perceptions on missing data**: by focusing on how users perceive and deal with missing data in PI systems, my work adds depth to the understanding of user needs and behaviors. This understanding is crucial for designing systems that are more user-friendly and can effectively support users in their self-tracking efforts and aligns with the call for more user-centric research in PI by Li et.al. [62] and Epstein et.al. [41], who emphasize understanding personal tracking behaviors and challenges.
- **Exploration of methods to address missing data**: The exploration and comparison of different methods to resolve missing data enrich the toolkit available for PI system designers. By examining the effectiveness and user preferences regarding these methods, my work guides the development of more effective solutions for data gaps identified by Rooksby et.al. [84], who discuss the everyday challenges of tracking and the need for more adaptable systems. My exploration adds practical solutions for one of these challenges.
- **Design implications for future PI systems**: The design implications derived from the studies serve as a valuable framework for future development in the PI field. By outlining how systems can better mediate and explain missing data, my thesis provides a blueprint for creating more engaging, understandable, and trustworthy PI tools. This approach is in line with the design considerations suggested by Consolvo et.al. [25], who highlight the importance of designing with the user in mind.
- **Empirical Insights into User Trust and System Transparency**: By demonstrating that user trust significantly relies on their understanding of how systems estimate missing data, the thesis underscores the necessity for transparency and explainability in design. These insights can lead to the development of PI systems that are not only more user-friendly but also foster greater

confidence among users in the accuracy and reliability of the data presented. This supports the arguments by Ribeiro et al. [82] and Lipton [66], who advocate for explainable models that users can trust and understand.

This thesis work enriches the PI field by trying to shed light on a critical yet underexplored issue, applying a user-centered lens to the challenge of missing data. These contributions not only advance our understanding of PI systems and their users but also lay the groundwork for future research and development in creating more robust, user-friendly, and trustworthy tools.

8.0 Limitations

[Methodological Limitations:]

Methodological limitations in this research are related to the size and demographics of the sample. In the semi-structured interview study (see Chapter 4 for details), we recruited 20 participants—10 males and 10 females from across the US. The recruitment was conducted through Prolific, an online research platform facilitating participant recruitment and management. While this sample size is adequate for a research user study, it may not be extensive enough for the entire user base. From this study, we identified two groups: *maintainers* and *trainees*(shown in Section 4.2.2). And their requirements are different. Trainees were inclined to seek specific insights and compare precise data across various timeframes. Maintainers focused on obtaining a summary of insights that encompass a wide range of past activities. However, with more participants, other groups are likely to emerge. The classification into *maintainers* and *trainees* is based on how they use their PI devices. If we categorize the groups based on other factors such as gender, race, age, nationality, family status, and marital status, different usage patterns may emerge. The findings are based on the qualitative research method. The findings might have differed if we had conducted a statistical analysis where we had access to logs of how people use these PI devices.

In the iterative user-centered design study (see Chapter 6 for details), we recruited 14 participants who claim to be maintainers, each of whom had four sessions with us. The recruitment was conducted through Pitt+Me, a service provided by the Clinical and Translational Science Institute at the University of Pittsburgh. In this study, we have 2 male and 12 female participants. The imbalance in the gender of our participants might influence the implications we derived from the study. And all of them are from Pittsburgh, so they might not represent the entire user base. We specifically recruited individuals using Fitbit devices for the iterative user-centered design study. Whether the specific types of devices among maintainers significantly impact their requirements or their usage behaviors are similar to those of other wearable devices remains uncertain.

Most participants in our study were office workers, students, and engineers, with none engaged in hard labor. Although we tried to recruit participants more aggressively, due to time constraints, we had biases in the recruited population. If we had more time, things would be done differently.

We recognize limitations in the diversity of the recruited population, and there may be additional behaviors to uncover by including individuals from various occupations and diverse backgrounds. However, such concerns are common in user studies. A thorough analysis of this data type, resembling a vendor's approach, where actual data usage is represented, can be considered to mitigate this. The orientation of our study is exploratory, and extensive analyses like A/B testing may not align with our research goals. The participatory design method was chosen to explore the design space of systems mediating missing data, as it allows us to delve into user needs more deeply than other methods. While data examination can provide insights into feature relevance, interviews enable a direct and profound understanding of user needs and the desired features for the system.

[Algorithmic Limitations:]

From the semi-structured study (see Chapter 4 for details), we found maintainers consistently engaged with the tools during data collection. However, in the iterative user-centered design study (see Chapter 6 for details) conducted in a lab setting, this aspect was not tested at the time of data collection but rather later. We utilized participants' Fitbit data for the past two weeks, intentionally removed some existing data, and asked them to use the system to estimate the data we removed. Although derived from the semi-structured study, the representation of this hypothetical scenario where missing data occurs may have unknown impacts. In the semi-structured study, some participants reported that they would estimate the data based on how their body feels, which normally happens shortly after their activity ends. However, in the iterative user-centered design study, the estimation process started after a few days of the activity (depending on which day we removed). We received comments from participants saying they no longer remember the activity or the day. Some design implications included helping participants recall the missing day. Still, the findings might differ if the study was conducted shortly after they had the activity, like after a few hours or less than a week.

We implemented three estimation methods in the iterative user-centered design study: algorithmic, event-based, and manual input (see Chapter 5 for details). The algorithmic approach utilized a wizard-of-oz method, presenting estimated data with a precision equivalent to what the state-of-the-art can achieve without implementing the model. We could build the prediction model and conduct experiments on that. our primary goal was to explore the design space of systems assisting end-users in estimating missing data. Thus, the chosen method is sufficient for this exploratory study.

[Data Limitations:]

One notable limitation in this thesis is related to data availability and quality. In the early work (see Chapter 3 for more details), we utilized a preexisting dataset to explore existing methods dealing with missing data. In this study, part of the challenge is distinguishing between when data is missing and when there's simply no data available. Existing imputation models perform well in estimating overall missing data (e.g., daily step data) but encounter challenges when dealing with time-sequenced data, such as running mileage starting at 2 pm on a specific day.

While the dataset is comprehensive, it carries inherent limitations. Firstly, the dataset lacked granularity in certain key variables, hampering a deeper understanding of the nuances of the phenomena under investigation. For instance, only total distance data is available, with no hourly or minute-level distance data (refer to Table 4), making it impossible to identify distance per activity. Secondly, due to privacy regulations governing the dataset, our study relied on aggregated data (as noted in the dataset description in [73]), limiting our ability to analyze individual-level variations. Thirdly, temporal gaps in the dataset were identified, impacting the establishment of continuous trends over extended periods. Although efforts were made to mitigate these issues through data cleaning and validation, these limitations might affect the generalizability of the findings.

[Privacy Limitations:]

One notable privacy concern related to participants referencing their friends' data to inform their own estimates when dealing with missing data. In the semi-structured study (see Section 4.2.3), we found that 20% of maintainers and 30% of trainees indicated they would refer to their friends' data for estimating their PI data when it was unavailable. However, due to privacy limitations, we don't have access to their friends' data, and the participants in our study are not friends with each other. Although our participatory study (detailed in Section 6.2) did not reveal a need for participants to refer to friends' data, feedback from the semi-structured user study suggested that participants might consider using their friends' data. There might be new implications if we had incorporated their friends into the study. For instance, individuals may not want an application to talk with their friends on their behalf or only want the app to contact their friends under certain circumstances. New implications might be discovered as trade-offs to balance privacy with automating the estimation process.

The current design implications specifically suggest that the system should learn more about people's daily habits (see Section 7.1) and routines (see Section 7.2) to assist them in recalling and estimating missing data. However, these features may raise privacy concerns and may not be actionable if users do not wish to share their personal data with the system. For instance, users might find it obtrusive when the system prompts them for location confirmation upon detecting new locations. The design implication also suggests that the question prompts should be personalized to be relevant to each user, which might also raise privacy concerns if users are unwilling to share certain levels of detail about their personal information. Additionally, these preferences might be dynamic. For example, someone with a serious health condition might find it more critical to allow the app to access more personal information, including medical records. However, someone who uses the application for general fitness might prefer to share more generalized private information. These diverse preferences will also influence the design of the system.

[Other:]

As described in Section 4.2, maintainers are a different set that, as observed through their engagement with the system (as shown in Section 4.2.2), tend to reflect on their past behaviors toward their goals while staying focused on the present. In contrast, trainees often seek to comprehend their past activities and make predictions for the future. Subsequent studies (outlined in Section 6.2) revealed two distinct personalities within the maintainer group: wellness-oriented and performance-oriented, each with unique requirements. Wellness-oriented individuals have desired outcomes or achievements without necessarily tying them to numerical values. On the other hand, performance-oriented individuals set clear, precise goals that can be assessed using quantifiable metrics. These individuals may also follow either predefined or flexible exercise routines. Based on personality traits and exercise routines, four distinct groups were identified: group SW (see Section 6.2.1), group SP (see Section 6.2.2), group UW (see Section 6.2.3), and group UP (see Section 6.2.4).

In this study, we had 14 participants. While we gathered diverse feedback to drive implications from them, recruiting more participants could potentially reveal additional groups and implications. Our goal in this thesis is not to exhaustively study all factors but to identify high-level differences that can inform the design of systems in this domain. The present findings remain valid with how the current study preceded, but conducting a larger-scale study might lead to new insights.

The system is specifically designed for maintainers, and throughout the iterative user-centered design study, we exclusively involved participants identified as maintainers, not trainees. Some groups, such as SW (see Section 6.2.1), exhibited behaviors similar to trainees, seeking specific insights and comparing precise data across various timeframes, although to a lesser extent than trainees but with comparable behaviors. While the lessons learned from the maintainers may transfer to trainees, the absence of trainees in its study introduces uncertainty regarding the presence of similar personalities in that group and the identification of comparable groups.

Throughout the iterative study, some participants expressed skepticism about their Fitbit device. For instance, one participant believed that the calories provided by Fitbit were an average estimation, and as he did not see himself as an average person, he distrusted the calorie feature. Other participants questioned how the Fitbit device calculated the calories, leading them to avoid using this feature. This lack of confidence in their devices might have unknown effects on how they perceive the estimated data. However, this limitation is common in technology in general; most apps are not fine-tuned for the entire population and are often designed for specific groups, leaving some users underserved. Therefore, while our participants reported this concern, it reflects a widespread limitation in many research studies.

9.0 Future directions

One valuable avenue for future exploration is the delicate balance between privacy considerations and the estimation of missing data. The implications of this work (see Chapter 7) highlight that methods employed in the system call for an understanding of users' daily habits (see Section 7.1) and routines (see Section 7.2), especially in the context of automating the process. The system collects comprehensive daily habits to construct a personalized and adaptive estimation model, capturing user routine changes for precision and relevance of the estimation process. Participants exhibit varying preferences, some favoring highly automated systems and others preferring less automation. Future research should delve into how participants' privacy concerns impact the design of such systems, seeking to strike the right balance between maintaining privacy and providing a personalized estimating model. Additionally, studies exploring alternative methods of collecting personal information, such as through ubiquitous sensors, are crucial. Although this study suggests using prompts, acknowledging that some participants might find them cumbersome, the potential preference for this method due to privacy concerns remains an interesting field to explore. When building the adaptive model using daily habits, it's essential to determine which habits to include in the models. Similarly, when using prompts to acquire personal information, it's crucial to identify the types of questions that can be asked and those that users may find sensitive and should be excluded. The study also suggests that the algorithmic model should incorporate user prompts for personalization, but the number, frequency, and type of prompts are under-explored. New studies should be conducted to address these questions. In the semi-structured user study, some participants indicated that estimating calories burned is acceptable, but not estimating heart rate (see Section 4.2.3). Hence, it's crucial to explore what types of data users feel comfortable estimating when missing and identify the data they find uncomfortable estimating.

In addition, conducting dedicated studies specifically focused on trainees could be valuable. The nature of our thesis study is exploratory, revealing higher-level distinctions between trainees and maintainers. Further studies could be essential to gain a more nuanced understanding of trainees. As highlighted in the limitations section (see Chapter 8), the implications derived in this work are guided by maintainers. Given the divergent usage behaviors between trainees and maintainers, exploring new implications tailored to trainees becomes applicable. Moreover, trainees may exhibit unique personalities and exercise habits distinct from maintainers, adding an additional layer of complexity to the design space. Therefore, exploring the design considerations for trainees promises to be an interesting avenue.

Another promising avenue for exploration in the context of Personal Informatics (PI) data is quantitative analysis. While this work primarily relies on qualitative analysis, a notable gap exists in quantitative approaches. To the best of our knowledge, sophisticated neural networks for data imputation are predominantly applied to large datasets rather than PI data. Developing imputation models tailored to scattered and smaller datasets could significantly contribute to addressing gaps in estimating missing data. Additionally, the implications drawn from this thesis advocate for leveraging users' daily habits and routines to construct personalized and adaptive estimation models, as outlined in Sections 7.1 and 7.2. After constructing a more dedicated system aligned with these implications, quantitative analyses can be conducted to refine the designs. For instance, the Fitbit company could implement a small feature that mediates missing data, allowing a portion of its users to use its app with this feature and another portion without it. With a larger user base like Fitbit's, they are able to overcome the limitations we encounter in the existing approach.

Bibliography

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–18, 2018.

[2] Chadia Abras, Diane Maloney-Krichmar, Jenny Preece, et al. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456, 2004.

[3] Amina Adadi and Mohammed Berrada. Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, 6:52138–52160, 2018.

[4] Giuseppe Agapito, Barbara Calabrese, Pietro Hiram Guzzi, Mario Cannataro, Mariadelina Simeoni, Ilaria Caré, Theodora Lamprinoudi, Giorgio Fuiano, and Arturo Pujia. Dietos: A recommender system for adaptive diet monitoring and personalized food suggestion. In *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 1–8. IEEE, 2016.

[5] Luis Alvarez. What does being physically fit mean? https://www.medicalnewstoday.com/articles/7181.

[6] Jorge Alvarez-Lozano, Venet Osmani, Oscar Mayora, Mads Frost, Jakob Bardram, Maria Faurholt-Jepsen, and Lars Vedel Kessing. Tell me your apps and i will tell you your mood: correlation of apps usage with bipolar disorder state. In *Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments*, pages 1–7, 2014.

[7] JULEN ARANGUREN. Fitbit fitness tracker data. https://www.kaggle.com/code/julenaranguren/bellabeat-case-study.

[8] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.

[9] Christiane Attig and Thomas Franke. Abandonment of personal quantification: A review and empirical study investigating reasons for wearable activity tracking attrition. *Computers in Human Behavior*, 102:223–237, 2020.

[10] CQUniversity Australia. Counting your steps. https://www.10000steps.org.au/articles/healthy-lifestyles/counting-steps/.

[11] Eric PS Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. Reviewing reflection: on the use of reflection in interactive system design. In *Proceedings of the 2014 conference on Designing interactive systems*, pages 93–102, 2014.

[12] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(5):1–27, 2013.

[13] Nerissa Campbell, Stefanie De Jesus, and Harry Prapavessis. *Physical Fitness*, pages 1486–1489. Springer New York, New York, NY, 2013.

[14] Victoria Catterall-Decalmer. How much time do americans spend on their health and fitness. https://us.myprotein.com/thezone/training/american-health-fitness-survey/.

[15] ANASTASIIA CHEBOTINA. Fitbit fitness tracker data. https://www.kaggle.com/code/chebotinaa/bellabeat-case-study-with-r/notebook.

[16] Yu Chen, Jiyong Zhang, and Pearl Pu. Exploring social accountability for pervasive fitness apps. In *Proceedings of the eighth international conference on mobile ubiquitous computing, systems, services and technologies*, pages 221–226, 2014.

[17] Eun Kyoung Choe, Bongshin Lee, et al. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications*, 35(4):28–37, 2015.

[18] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A Kientz. Sleeptight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 121–132, 2015.

[19] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pages 173–182, 2017.

[20] Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt, and Julie A Kientz. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1143–1152, 2014.

[21] Woohyeok Choi, Sangkeun Park, Duyeon Kim, Youn-kyung Lim, and Uichin Lee. Multi-stage receptivity model for mobile just-in-time health intervention. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(2):1–26, 2019.

[22] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A Kientz, and Sean A Munson. 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*, pages 770–786, 2016.

[23] CMS.com. Nhe fact sheet. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.

[24] Physical Activity Guidelines Advisory Committee. Physical activity guidelines advisory committee scientific report. washington, dc: Us department of health and human services; 2018. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://health.gov/sites/default/files/2019-09/pag$_a dvisory_c ommittee_r eport.pdf$.

[25] Sunny Consolvo, Predrag Klasnja, David W McDonald, Daniel Avrahami, Jon Froehlich, Louis LeGrand, Ryan Libby, Keith Mosher, and James A Landay. Flowers or a robot army? encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 54–63, 2008.

[26] Sunny Consolvo, David W McDonald, and James A Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 405–414, 2009.

[27] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1797–1806, 2008.

[28] Lynn Coorevits and Tanguy Coenen. The rise and fall of wearable fitness trackers. In *Academy of Management*, 2016.

[29] Andrea Cuttone, Sune Lehmann, and Jakob Eg Larsen. A mobile personal informatics system with interactive visualizations of mobility and social interactions. In *Proceedings of the 1st ACM international workshop on Personal data meets distributed multimedia*, pages 27–30, 2013.

[30] Andrea Cuttone, Michael Kai Petersen, and Jakob Eg Larsen. Four data visualization heuristics to facilitate reflection in personal informatics. In *Universal Access in Human-Computer Interaction. Design for All and Accessibility Practice: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part IV 8*, pages 541–552. Springer, 2014.

[31] Jacqueline DeMarco. Nearly 70% of americans would wear a fitness tracker/smartwatch for discounted health insurance. https://www.valuepenguin.com/fitness-tracker-smartwatch-health-survey.

[32] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, and Dan Dixon. Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems*, pages 2425–2428. 2011.

[33] Derek Doran, Sarah Schulz, and Tarek R Besold. What does explainable ai really mean? a new conceptualization of perspectives. *arXiv preprint arXiv:1710.00794*, 2017.

[34] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

[35] Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, et al. Explainable ai (xai): Core ideas, techniques, and solutions. *ACM Computing Surveys*, 55(9):1–33, 2023.

[36] Tlamelo Emmanuel, Thabiso Maupong, Dimane Mpoeleng, Thabo Semong, Banyatsang Mphago, and Oteng Tabona. A survey on missing data in machine learning. *Journal of Big Data*, 8(1):1–37, 2021.

[37] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 conference on Designing interactive systems*, pages 667–676, 2014.

[38] Daniel A Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A Munson. Beyond abandonment to next steps: understanding and designing for life after personal informatics tool use. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1109–1113, 2016.

[39] Daniel A Epstein, Bradley H Jacobson, Elizabeth Bales, David W McDonald, and Sean A Munson. From“ nobody cares” to“ way to go!” a design framework for social sharing in personal informatics. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1622–1636, 2015.

[40] Daniel A Epstein, Nicole B Lee, Jennifer H Kang, Elena Agapie, Jessica Schroeder, Laura R Pina, James Fogarty, Julie A Kientz, and Sean Munson. Examining menstrual tracking to inform the design of personal informatics tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6876–6888, 2017.

[41] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 731–742, 2015.

[42] Kelly R Evenson, Michelle M Goto, and Robert D Furberg. Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1):1–22, 2015.

[43] Ahmed Fadhil. Beyond technical motives: Perceived user behavior in abandoning wearable health & wellness trackers. *arXiv preprint arXiv:1904.07986*, 2019.

[44] Ahmed Fadhil. Different stages of wearable health tracking adoption & abandonment: A survey study and analysis. *arXiv preprint arXiv:1904.13226*, 2019.

[45] Centers for Disease Control and Prevention. Atlas. ti 8 windows-inter-coder chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://downloads.atlasti.com/docs/ica$_{a}8_{w}in_{e}n.pdf$.

[46] Centers for Disease Control and Prevention. Physical activity and health. https://www.cdc.gov/physicalactivity/basics/pa-health/index.htm. accessed february 22, 2023.

[47] Sarah Fox, Noura Howell, Richmond Wong, and Franchesca Spektor. Vivewell: Speculating near-future menstrual tracking through current data practices. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, pages 541–552, 2019.

[48] Thomas Fritz, Elaine M Huang, Gail C Murphy, and Thomas Zimmermann. Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 487–496, 2014.

[49] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. How do we engage with activity trackers? a longitudinal study of habito. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 1305–1316, 2015.

[50] John W Graham and John W Graham. Analysis of missing data. *Missing data: Analysis and design*, pages 47–69, 2012.

[51] World health organization. Mental health. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response.

[52] Cheng-Kang Hsieh, Hongsuda Tangmunarunkit, Faisal Alquaddoomi, John Jenkins, Jinha Kang, Cameron Ketcham, Brent Longstaff, Joshua Selsky, Betta Dawson, Dallas Swendeman, et al. Lifestreams: A modular sense-making toolset for identifying important patterns from everyday life. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, pages 1–13, 2013.

[53] Md Hamidul Huque, John B Carlin, Julie A Simpson, and Katherine J Lee. A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC medical research methodology*, 18(1):1–16, 2018.

[54] Nassim Jafarinaimi, Jodi Forlizzi, Amy Hurst, and John Zimmerman. Breakaway: an ambient display designed to change human behavior. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1945–1948, 2005.

[55] Daoyan Jin, Hallgeir Halvari, Natalia Maehle, and Anja H Olafsen. Self-tracking behaviour in physical activity: a systematic review of drivers and outcomes of fitness tracking. *Behaviour & Information Technology*, 41(2):242–261, 2022.

[56] Hyun Kang. The prevention and handling of the missing data. *Korean journal of anesthesiology*, 64(5):402–406, 2013.

[57] Robert W Krause, Mark Huisman, Christian Steglich, and Tom AB Snijders. Missing network data a comparison of different imputation methods. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 159–163. IEEE, 2018.

[58] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. Tell me more? the effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 1–10, 2012.

[59] MACARENA LACASA. Fitbit fitness tracker data. https://www.kaggle.com/code/macarenalacasa/capstone-case-study-bellabeat.

[60] Federica Laricchia. Number of wearable device users in the u.s. 2014-2022. https://www.statista.com/statistics/543070/number-of-wearable-users-in-the-us/.

[61] Amanda Lazar, Christian Koehler, Theresa Jean Tanenbaum, and David H Nguyen. Why we use and abandon smart devices. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 635–646, 2015.

[62] Ian Li, Anind Dey, and Jodi Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 557–566, 2010.

[63] Ian Li, Anind K Dey, and Jodi Forlizzi. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 405–414, 2011.

[64] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. Fish'n'steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*, pages 261–278. Springer, 2006.

[65] Suwen Lin, Xian Wu, Gonzalo Martinez, and Nitesh V Chawla. Filling missing values on wearable-sensory time series data. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, pages 46–54. SIAM, 2020.

[66] Zachary C Lipton. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57, 2018.

[67] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

[68] Alessandro Marcengo and Amon Rapp. Visualization of human behavior data: the quantified self. In *Big Data: Concepts, Methodologies, Tools, and Applications*, pages 1582–1612. IGI Global, 2016.

[69] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.

[70] Jun-Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I Hong. Toss'n'turn: smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 477–486, 2014.

[71] Varun Mishra, Florian Künzler, Jan-Niklas Kramer, Elgar Fleisch, Tobias Kowatsch, and David Kotz. Detecting receptivity for mhealth interventions in the natural environment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2):1–24, 2021.

[72] Jane E Myers, Thomas J Sweeney, and J Melvin Witmer. The wheel of wellness counseling for wellness: A holistic model for treatment planning. *Journal of counseling & development*, 78(3):251–266, 2000.

[73] MÖBIUS. Fitbit fitness tracker data. https://www.kaggle.com/datasets/arashnic/fitbit?resource=download.

[74] Erdman B Palmore. How to live longer and like it1. *Journal of Applied Gerontology*, 4(2):1–8, 1985.

[75] Maria Pampaka, Graeme Hutcheson, and Julian Williams. Handling missing data: analysis of a challenging data set using multiple imputation. *International Journal of Research & Method in Education*, 39(1):19–37, 2016.

[76] pewResearch. Americans views on mobile devices and health. http://www.pewinternet.org/2018/08/22/americans-views-on-mobile-devices-and-health-2/.

[77] PrincetonUMatter. Princeton umatter wellness self-assessment. https://umatter.princeton.edu/action/caring-yourself/wellness-wheel-assessmentassessment.

[78] PrincetonUMatterQuestionnaire. Princeton umatter wellness self-assessment questionnaire. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://umatter.princeton.edu/sites/g/files/toruqf2181/files umatter-wellness-self-assessment.pdf.

[79] J Ross Quinlan. Unknown attribute values in induction. In *Proceedings of the sixth international workshop on machine learning*, pages 164–168. Elsevier, 1989.

[80] Amon Rapp and Federica Cena. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies*, 94:1–17, 2016.

[81] Amon Rapp, Alessandro Marcengo, Luca Buriano, Giancarlo Ruffo, Mirko Lai, and Federica Cena. Designing a personal informatics system for users without experience in self-tracking: a case study. *Behaviour & Information Technology*, 37(4):335–366, 2018.

[82] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[83] Diana Rodriguez. Why exercise boosts mood and energy. https://www.everydayhealth.com/fitness/workouts/boost-your-energy-level-with-exercise.aspx.

[84] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. Personal tracking as lived informatics. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1163–1172, 2014.

[85] Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[86] Herman Saksono, Carmen Castaneda-Sceppa, Jessica A Hoffman, Magy Seif El-Nasr, and Andrea Parker. Storymap: Using social modeling and self-modeling to support physical activity among families of low-ses backgrounds. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

[87] Benjamin Saunders, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, 52(4):1893–1907, 2018.

[88] Cliff Scott and Melissa Medaugh. Axial coding. *The international encyclopedia of communication research methods*, 10:9781118901731, 2017.

[89] Hamza Sellak and Marthie Grobler. mhealth4u: designing for health and wellbeing self-management. In *2020 35th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*, pages 41–46. IEEE, 2020.

[90] Jasmit S Shah, Shesh N Rai, Andrew P DeFilippis, Bradford G Hill, Aruni Bhatnagar, and Guy N Brock. Distribution based nearest neighbor imputation for truncated high dimensional data with applications to pre-clinical and clinical metabolomics studies. *BMC bioinformatics*, 18(1):1–13, 2017.

[91] Moushumi Sharmin, Andrew Raij, David Epstien, Inbal Nahum-Shani, J Gayle Beck, Sudip Vhaduri, Kenzie Preston, and Santosh Kumar. Visualization of time-series sensor data to inform

the design of just-in-time adaptive stress interventions. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 505–516, 2015.

[92] Tacksoo Shin, Mark L Davison, and Jeffrey D Long. Maximum likelihood versus multiple imputation for missing data in small longitudinal samples with nonnormality. *Psychological methods*, 22(3):426, 2017.

[93] SurveyMonkey. Demographic survey question guide: Tips, examples, best practices. https://www.surveymonkey.com/mp/gathering-demographic-information-from-surveys/.

[94] Melanie Swan. Sensor mania! the internet of things, wearable computing, objective metrics, and the quantified self 2.0. *Journal of Sensor and Actuator networks*, 1(3):217–253, 2012.

[95] Amira Tlija, K Wegrzyn-Wolska, and Dan Istrate. Missing-data imputation using wearable sensors in heart rate variability. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, (2), 2020.

[96] Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B Altman. Missing value estimation methods for dna microarrays. *Bioinformatics*, 17(6):520–525, 2001.

[97] Stef Van Buuren. *Flexible imputation of missing data*. CRC press, 2018.

[98] Greg Walsh and Jennifer Golbeck. Stepcity: a preliminary investigation of a personal informatics-based social game on behavior change. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2371–2376. 2014.

[99] Darren ER Warburton, Crystal Whitney Nicol, and Shannon SD Bredin. Health benefits of physical activity: the evidence. *Cmaj*, 174(6):801–809, 2006.

[100] Wikipedia. Apple watch. https://en.wikipedia.org/wiki/apple$_{w}atch$.

[101] Wikipedia. Fitbit. https://en.wikipedia.org/wiki/fitbit.

[102] Wikipedia. Nike+ipod. https://en.wikipedia.org/wiki/nike

[103] Wikipedia. Nvivo. https://en.wikipedia.org/wiki/nvivo.

[104] Wikipedia. Quantified self. https://en.wikipedia.org/wiki/quantified$_{s}elf$.

[105] Xian Wu, Stephen Mattingly, Shayan Mirjafari, Chao Huang, and Nitesh V Chawla. Personalized imputation on wearable-sensory time series via knowledge transfer. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1625–1634, 2020.