SMART MEASURES OF SOCIAL WITHDRAWAL IN PEOPLE WITH PARKINSON'S DISEASE

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF SCIENCE AND ENGINEERING

2022

Heng Zhang

School of Engineering, Department of Computer Science

Contents

Word Count: 53803

List of Tables

List of Figures

Abstract

This work aims to explore social withdrawal in people with Parkinson's disease (PD), an incurable, neurodegenerative disease that impacts 1% of people over the age of 60 around the world. PD causes a wide range of motor and non-motor symptoms that significantly influence people's quality of life (QoL). Both motor and non-motor symptoms could cause social withdrawal, and social wellbeing plays a critical role in patients' QoL. Therefore, social withdrawal could be a significant consequence of disease progression and deterioration of QoL.

For a disease without a complete cure, the principal aim of the treatment is to improve the QoL of patients. The measurement of Parkinson's progression is the prerequisite of appropriate treatment. However, the clinical assessment of Parkinson's is usually conducted every six months and is based on a snapshot of symptoms which might vary over time. These assessments are done by patients or experts and could be biased by memory or experience. Patients are also aware they are being assessed, which may introduce a Hawthorne effect. Thus, continuous, unobtrusive, and objective measurement is the ambition of Parkinson's monitoring. The smartphone is a popular digital device, and it consumes a significant amount of time for personal, social communications. With embedded sensors, the smartphone can even infer social interactions external to it. Previous studies have confirmed its feasibility for observing people's behaviour without disruption. So, it is a promising tool for unobtrusively tracking social activities, and it fulfils the purpose of a novel monitoring method.

Therefore, we initiated a year-long longitudinal study to explore social withdrawal in PD patients. A monitoring application was installed on participants' smartphones to capture all nine potentially social-related data sources, 24 hours a day, seven days a week. Eight standardised clinical/psychological scales for measuring Parkinson's progression, QoL, social withdrawal, and related factors were conducted every two

months. Specifically designed diaries were also provided to participants to record their weekly QoL and level of social interaction. With participants joining and dropping out, eight participants finished the whole year of observation. As the continuous monitoring of the smartphone application, more than 10 million raw smartphone data points were obtained from these eight participants. Then twenty-two features were extracted from these raw data to establish personal understandings of the social behaviour of each participant.

The COVID-19 pandemic, which significantly impacted people's social lives, occurred during the experiment. But it also provided an opportunity to examine our approach to detecting severe social impact. With the confirmation from the interviews with participants, our method successfully reflected participants' conformance to the government's policies for reducing the transmission of COVID-19 and the intense social deviations caused. For the outcomes of the whole-year study, significant associations were found between clinical/psychological scales and at least one feature for each individual. Our model also achieved at least 0.6 R-squared in numerical prediction and 0.6 F1 scores in direction projection of all participants using multiple linear regression and Naive Bayes methods. Overall, we presented an approach that can adaptively learn the social behaviour of a particular individual and make predictions based on smartphone data. It also shows its strong potential as a reference for clinical/psychological standards. Future work can build upon our efforts to more comprehensive monitoring and a higher validity of the approach. The technique demonstrated in this work could also be applied in wider communities where the patients' social impact needs attention.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

To my supervisors, Simon Harper, Bijan Parsia, Ellen Polikoff, for being supreme mentors that led me the way to explore the edge of knowledge, I have benefited greatly from your teaching, and it will be remembered in my life.

To all participants involved in this study, you were always kind, patient and enthusiastic, which inspired us to overcome difficulties and work hard. I hope this study is a step toward in-depth Parkinson's understanding and better Parkinson's care.

To all the colleagues in the IAM lab, especially Julio and Ahmed, your knowledge and help gave me great support, which saved me a lot of detours.

To Haoruo Zhao, Sen Zheng, Mimi Twaijri and other friends in Manchester, you made this journal magnificent. The time with you was the most enjoyable part of my whole PhD life. Special thanks to Mimi. Without her, I could not complete the participants' interviews for the study.

特别感谢我的父母张建华和吴建荣，他们是这个项目的实际出资人，感谢他们一直以来的支持和供养。

最后感谢下自己，误打误撞的开启了博士生涯，也算好好体验了一番。用张若虚《春江花月夜》其中的三局结尾吧，它们也被用在了高中三年每年的年度总结里，正好也是对这几年的概括
江畔何人初见月？江月何年初照人？
人生代代无穷已，江月年年望相似
不知江月待何人，但见长江送流水

Chapter 1

Introduction

This work explores the use of smartphone digital phenotyping in monitoring the social behaviours of individuals. We focus on the particular behaviour of social withdrawal in people with Parkinson's disease (PD), as it is a significant consequence of disease progression and has an impact on patients' quality of life (QoL). Digital phenotyping refers to the technology using digital devices to measure people's daily behaviours continuously without asking questions [164]. This work analyses data from a person's smartphone, including social (phone calls, text messages, social media, conversations) and physical (Global Positioning System [GPS]) data to infer a social activity level. It opens another door to understanding the social behaviours of a particular population. Rather than utilising snapshot and easily biased questionnaires, smartphone monitoring can provide continuous and objective data. More detailed and individualised understanding could be built on these data. Furthermore, this work indicates that personalised health services and the social wellbeing of wider communities could be promoted by digital phenotyping.

Human beings are social creatures. We all live in a collaborative environment and interact with each other to survive and thrive by nature [58]. A social life is essential for everyday life, and social activity plays a significant role in our mental and physical health. A recent meta-analysis showed that more socially isolated people have a higher mortality rate than the general population [96]. The health status of those who are socially isolated is lower than the population of the same age who are not socially isolated [92]. In addition, one of the crucial risks of depression social relationships of poor quality [225]. On the contrary, positive social activity benefits both individuals and society. A higher degree of social integration is associated with healthy systolic blood pressure, body mass index, and waist circumference, and lower inflammation markers [251]. Better cognitive function often exists with people who have more social support [198]. Moreover, the productivity of workers could be increased by saving time utilising effective social communications [7].

Therefore, understanding social behaviour is meaningful and beneficial, and social behaviour research grasped attention decades ago [91]. Typically, those experiments were conducted in a controlled setting like a laboratory, and a stimulus was initiated towards participants. Then the reactions of participants were observed and measured [235]. The form of the measurement includes general scales, specially designed questionnaires, voice or video recordings, and human observations [167]. Nevertheless, there are issues with these forms of measurement. Questionnaires rely on people's memory, which could introduce recall bias. In controlled settings, the Hawthorne effect could appear [201]. Participants may be aware they are being monitored and therefore respond differently in their daily lives. More importantly, the cost of equipment and time of traditional research settings restrict the scale of experiments. It could be challenging to redo the whole procedure several times to achieve longitudinal results [80].

Passive sensing, which requires minimum interaction with participants, is a promising alternative. Data can be collected continuously in situ without observers [207]. The unobtrusiveness and ubiquity of passive sensing also make longitudinal studies feasible. It particularly benefits sensitive social behaviour research on mental health [43]. To accomplish these visions, dedicated devices also were designed. Electronically activated recorders can be applied to sample the sounds in participants' surroundings to infer social activities [142]. Badges equipped with radio frequency identification devices are treated as beacons to detect face-to-face interaction between participants wearing them [9]. Wearable sensors, including smartwatches and wristbands, also extract social behavioural cues and even have specialised sensors to monitor variables such as heart rate and sleep stages. But these sensors usually need another device, such as a smartphone, to transmit the data. Compared with smartphones, the penetration rate of wearable sensors is still low [214]. So, participants who usually do not wear a smartwatch or wristband may feel aware of being monitored, if it is required to be worn for the experiment[192].

Smartphones have become a popular personal digital device, with 87% of UK adults owning one in 2020. Ownership reaches 70% in people over age 55 [162]. Social sensing by smartphones could be less intrusive than by any other device. Smartphones also play an important role in social interactions. UK adults spend more than two hours a day on average on their smartphones [68]. Emails, messages, and social media are three of the most frequent smartphone activities [219]. Therefore, the smartphone could be an effective way to study social behaviours. With dedicated applications, social interaction on the smartphone, such as calls, messages, emails, and social media activity, could be captured without burden. Moreover, environmental sensors are embedded in almost every smartphone on the market that could enable the monitoring of social interactions, not only on the phone but also in the physical world. For example, the microphone could help detect whether the user is engaged in a conversation. Social contexts like locations could also be determined from GPS. In addition, customised applications can be built on smartphones, and data could be stored, transmitted, and processed efficiently on board, which could be convenient for deployment.

Parkinson's disease is a long-term neurological disease. Its prevalence increases with age, and 1% of people over the age of 60 around the world are affected [180]. The typical PD symptoms are movement disorders, including tremors, rigidity, slow movement, and walking difficulties. But other non-motor symptoms, such as cognitive problems, sleep disturbance, apathy, depression, and autonomic dysfunction, can also be triggered by PD [174]. Social withdrawal is when people gradually lose connection with other individuals and society. All these symptoms could cause the social withdrawal of the participants in this study. Motor symptoms may reduce their chance of going out and having social activities. Participants may have social difficulties, lose social confidence, and become socially anxious due to the impact of non-motor symptoms [206].

Currently, there is no cure for PD, and all treatments aim to alleviate symptoms and improve QoL. Monitoring the progression of PD is the prerequisite of the appropriate treatment. Notably, PD advancement and symptoms are idiosyncratic [107], and participants have their own unique path and speed of disease evolution. This adds more difficulty to the treatment of the disease since doctors have to assess the comprehensive condition to make reasonable decisions on treatment. Since various PD symptoms can cause social withdrawal, the extent of social withdrawal could be a significant consequence of PD progression. PD patients' social behaviour also provides doctors with another perspective to understand the impact on QoL. So, more targeted treatment could be achieved since social factors have an essential role in QoL [246].

However, understanding the social withdrawal in PD is still at an early stage. Researchers rely on previous knowledge and participants' reports to comprehend the social symptoms of PD [178]. It could be subjective and short compared with the long time in which PD symptoms evolve. To the best of our knowledge, no existing studies have investigated the social withdrawal of PD patients in a long-term manner. We are the first to conduct a year-long longitudinal study to investigate the social withdrawal in Parkinson's disease using the digital phenotyping method. The results demonstrate that our method successfully tracks PD patients' social activity levels. The change in social habits due to severe external affairs – such as COVID-19 – was also revealed from our approach. These constructive results can also promote digital phenotyping in studying other long-term diseases that impact human behaviour. Further understanding and better care could be built on this novel knowledge. Moreover, it will benefit patients, carers, doctors, and researchers.

1.1 Research questions

The research questions of this PhD thesis are as follows:

RQ1: What data do we need to understand social behaviour, and how can we obtain them through digital phenotyping?

People have spent significant time on them for various kinds of social interactions, and smartphones can even infer social interactions external to the smartphones with the use of embedded sensors. There are existing preliminary smartphone sensing studies in which phone-based models outperform comparable demographic models [202]. It signifies the smartphone's potential to collect people's daily routines and behavioural characteristics. However, there is a gap between raw smartphone data and utilisable information. Unprocessed smartphone data could be massive and chaotic, and these data need to be processed in accordance with each data's properties to retrieve practical details. Therefore, we need to investigate the methods for extracting social events from the raw smartphone data to reconstruct the social behaviours of participants. Then, the individualised social behaviour model can be established from these social behaviours.

Through unobtrusive monitoring, behavioural knowledge obtained by digital phenotyping can contribute to a novel understanding of social behaviour.

RQ2: How can we recognise personal social engagement using digital phenotyping?

Practically, social withdrawal is defined as reduced social interactions. Not social behaviour itself but, rather, the change in behaviour related to time is the crucial variable we aim to measure. However, there are different quantifications of social interactions. With further combinations of these quantifications, more features could be generated to observe social engagement. These features could associate with insights into the deviation of social patterns. As a result, investigating feasible approaches to construct features from theoretical social impact can aid the transfer of this knowledge from its behavioural and psychological context to the realms of technology and informatics. It could help an in-depth understanding of human social behaviour and reveal possible health problems using unobtrusive approaches.

RQ3: How can we make connections between social behaviour collected by digital phenotyping and the clinical/psychological ground truth of Parkinson's patients?

Social behaviours can be described and measured in a variety of ways. Each communication channel has its own characteristics to be quantified. Moreover, people have their preferences for social contacts, such as channels of communication and types of contact. Not all of them are impacted equally by the disease. It also applies to Parkinson's induced social withdrawal. As a complex phenomenon, many factors like disease progression, QoL, and cognitive ability are involved. They may have different connections with any of the social factors. Studying them could provide a comprehensive understanding of social withdrawal. One of the aims of our study is to map smartphone data to existing clinical/psychological ground truth. Therefore, we need to discover which features have the most potential to reflect the social withdrawal, disease changes, and other related factors. This approach can also establish a personal understanding of social withdrawal towards PD.

RQ4: How can digital phenotyping be used to personalise social behaviour tracking in a more granular manner?

The typical PD scale is conducted every six months, and it is just a snapshot compared to daily fluctuated disease progression. QoL of patients is often neglected in the measurement. So, another ambition we aim to achieve is to provide more granular monitoring considering both progression and QoL, which is social withdrawal. Moreover, it is natural that individual differences influence people's social behaviour and perceived social interaction level. Some people may be extroverts who enjoy new social connections with others, whereas others may be satisfied with current relationships. A personal understanding is necessary both for disease and social behaviour. Therefore, we need an approach to learn the social patterns from an individual level. It should comprehend past social behaviours of that particular participant and establish a model for estimating the social interaction level from the self-report ground truth.

1.2 Contributions

The main contributions of this thesis are as follows:

1. **Knowledge that supports the understanding of social behaviour through digital phenotyping.**
   In Chapter 3, we completed a detailed systematic review of the digital phenotyping used to extract social behaviour passively from smartphones. Based on the findings from the systematic review, we then initiated a longitudinal study on actual PD patients to observe their social behaviours. Chapters 4 to 7 demonstrate various results on the way or at the end of the experiment. These results provide another perspective to understand social behaviour in PD patients. Particularly, the in-depth social impact of COVID-19 is discussed in Chapters 4 and 5. The relation between PD and smartphone social behaviour is illustrated in Chapters 6 and 7. This knowledge contributes to methods for collecting social behaviour through digital phenotyping and identifying the kind of data needed to understand human social behaviour (RQ1).

2. **Insights of social behaviour changes due to different reasons using digital phenotyping observations.**
   Theoretically, the progression of PD could cause reduced social interactions. However, unfortunately, the COVID-19 pandemic happened in the UK during our experiment period. The policies enforced by the government to reduce disease transmission severely damaged everyone's ordinary social life. We took advantage of this unique chance to observe the social behaviour changes of our participants. In Chapters 4 and 5, we creatively use features generated from

the smartphone data to observe the personal obedience of COVID-19-related restrictions. Their detailed communication changes and the insight of these transformations are also discussed. Apart from COVID-19, the association between PD ground truth and social behaviour changes is explained in Chapters 6 and 7. They were studied much more frequently than the typical six-month measurement interval. All this knowledge contributes to recognising personal social behaviour changes using digital phenotyping (RQ2) and personalising social behaviour tracking in a more granular manner (RQ4).

3. **An adaptive and personalised approach to measure social behaviour using digital phenotyping.** The aim of our project is to study social withdrawal in Parkinson's, so the key element we achieved is the approach to measure social behaviour from smartphone data. These measures were then connected with ground truth for every participant, which is discussed in Chapter 6. Our approach's adaptability to the data of each individual is demonstrated in Chapter 7. By selecting the most expressive social behaviour feature sets, we present a strategy for generating a personalised model from an overfitted set of social properties that forms the general approach. The model is specifically adapted to that person, but the approach to establishing the model can be utilised in wider health communities for digital phenotyping in social behaviour studies. All this knowledge contributes to making connections between social behaviour collected by digital phenotyping and the clinical/psychological ground truth of Parkinson's (RQ3) and personalising social behaviour tracking in a more granular manner (RQ4).

1.3 Thesis overview

This thesis is given in a journal/alternative format with the supervisory committee's approval from the Faculty of Science and Engineering. It means that the main chapters of this thesis (Chapters 3 to 7) are papers that have been published or presently are under review. The ability to read and interpret each paper separately drives our choice of journal format. On the other hand, chapters comprising these studies work together to contribute to the overall object of this thesis.

In this work, we explore the social withdrawal in PD patients by a longitudinal study using digital phenotyping. The basis of this work is twofold: 1) PD causes disruptions in patients' social functions by a slew of emotional and communicative changes [178], and 2) PD patients' behaviour can be tracked using an unobtrusive and personalised digital phenotyping methodology [94]. These previous works inspired us to monitor social behaviour in PD patients to promote their QoL and understanding of PD.

The major part of our work is a year-long longitudinal observation study of PD patients using smartphones. All chapters of this thesis are related to the design, implementation, or results of this study. Therefore, the thesis follows the timeline of the experiment, which is summarised in the following three phases: 1) theoretical preparation wherein we conducted a systematic review on passive smartphone social sensing, 2) severe change detection wherein participants' social impact by COVID-19 was detected, and wherein we innovatively created smartphone features according to COVID-19-related constraints to track individual adherence to these policies, and 3) analysis towards the ground truth wherein all-year data were gathered and processed. Both standardised clinical/psychological scales and more granular ground truth are considered.

1. **Theoretical preparation**
   To explore the social withdrawal in PD, the practical goal we aim to achieve is tracking the social behaviour of PD patients via smartphones. Although the smartphone has been employed to monitor participants' social behaviour, every study has its own understanding and arrangement for sensing strategy and subsequent analysis. The limitations and benefits of smartphone sensing that perhaps influence the study's outcome are also unrevealed. As a result, at the beginning of the study, we aim to make reasonable choices on the deployment of our monitoring system. The possible option is to discover studies using the same concept of this technology to investigate what they have accomplished and how the results were achieved. We culminated this with the paper titled *'Passive Social Sensing with Smartphone: A Systematic Review'*, which summarises all the procedures for unobtrusive digital phenotyping of social behaviour studies (Chapter 3). It strictly follows the state-of-art guideline of the systematic review process and comprehensively summarises each step of passive smartphone sensing. Sensing strategies, validation measures, data processing, feature extraction, and data analysis are all included in the systematic review. This work lays the foundation for which data we need in order to understand social behaviour and how we can obtain them through digital phenotyping (RQ1). Following the findings from the systematic review, we include in our design of the experiment the

possible social factors that can be captured by smartphones.

2. **Severe change detection**
After the theoretical preparation, all social-related data sources were selected from the findings from phase 1 to be tracked via smartphones. A monitoring application named AWARE [67] was chosen to record these data. After the ethics approval, the participant recruitment campaign started in September 2019, and the experiment was planned for a whole year. Unfortunately, the COVID-19 pandemic in the UK happened during the experiment period, and the government introduced a series of policies to reduce disease transmissions. People had to stay at home, maintain a social distance, and minimise face-to-face interactions with people outside their household. It inspired us to examine if the collected smartphone data can reveal participants' compliance with these policies. The results are published in a paper titled *'Digital Phenotypes for Understanding Individuals' Compliance with COVID-19 Policies and Personalised Nudges: Longitudinal Observational Study'* (Chapter 4). Moreover, these policies severely changed everyone's social life. We also compared the social behaviour before and during the pandemic to discover the impact of every communication channel. As vulnerable populations, these discoveries could provide personalised care during the pandemic. We summarised the result in the paper titled *'Monitoring Social Withdrawal with Smartphones in People with Parkinson's Disease and the Impact of COVID-19'* (Chapter 5). Possible negative health signals were found in participants, and they were explained via semi-structured interviews. All these works reflect the severe impact on participants' social behaviours, so it confirms that personal social behaviour changes can be detected using digital phenotyping (RQ2).

3. **Analysis towards the ground truth**
The experiment continued under the shadow of COVID-19. Apart from smartphone data, a set of scales, including PD progression, QoL, social withdrawal, and related factors, were conducted every two months. Specially designed diaries were provided to participants to record their weekly social extent and QoL. To keep the unobtrusiveness of the smartphone monitoring, participants did not do any tasks on their smartphone. So, these measures are the ground truth we depend on. The experiment ended in March 2021, and eight participants finished the whole year of observation. Different quantifications of social behaviour

were extracted as features from raw smartphone data. Then, correlation analyses were conducted towards the identified clinical/psychological scales. The details, methods, and results of this study are elaborated in the paper titled *'Exploring Social Withdrawal with Smartphones in People with Parkinson's Disease: A Longitudinal Study'* (Chapter 6). It connects social behaviour collected by digital phenotyping and the clinical/psychological ground truth of Parkinson's (RQ3). After successfully identifying the correlated social behaviour with two months' ground truth, we turned to more granular tracking of participants' social behaviour. Personalised models were constructed from the smartphone features towards weekly ratings. Both numerical and directional models were considered. The details of the models are summarised in *'Tracking Social Behaviour with Smartphones in People with Parkinson's: A Longitudinal Study'* (Chapter 7). It demonstrates in a more granular manner the capability of the approach of using digital phenotyping to personalise social behaviour tracking (RQ4). By solidifying the knowledge that has been produced, the thesis concludes in Chapter 8. The advantages and limitations of our work, and suggestions for future research prospects are also discussed in that chapter.

Chapter 2

Background

Using smartphones to measure social withdrawal in PD patients is the goal of this thesis. To establish the basic concept for achieving this goal, we present the theoretical foundations where the thesis was built. The introduction of PD, social withdrawal, and the relationship between PD and social withdrawal are provided in this chapter. In addition, the related work of digital phenotyping and its application in PD are also reviewed in this chapter.

2.1 Parkinson's disease

PD is a long-term neurodegenerative disease. It affects patients' central nervous system and motor system. It is particularly prevalent among elderly populations and is estimated to affect 1% of the population over 60 [190]. The symptoms of PD can be categorised into motor and non-motor symptoms. The most apparent PD symptoms are shaking, rigidity, bradykinesia and tremor, which may appear at an early stage. The onset of PD emerges gradually, and the non-motor symptoms become more evident as the disease progresses. These non-motor symptoms include behavioural and cognitive problems, typically depression, anxiety and apathy [87]. The cause of Parkinson's is still unknown. It is believed that both genetic and environmental factors contribute[105]. There is no cure for PD. All treatments aim to alleviate symptoms and improve patients' quality of life (QoL) [190]. Moreover, the progression and the symptoms of each PD patient are idiosyncratic [107]. Every one of them has unique patterns of the disease. A subset of a wide variety of different symptoms will be caused. Also, the development of symptoms is not linear but fluctuates, and different kinds of factors, such as medications, sleep and stress, may cause these fluctuations [71]. In summary, all these peculiarities cause difficulties in Parkinson's management and treatment.

A Parkinson's diagnosis requires four cardinal symptoms [101]: 1) bradykinesia, which is defined as the slow start of voluntary movement and gradual loss in speed and amplitude of alternating actions [105]; 2) rest tremors, usually rhythmic twitching movements at the distal part of an extremity with a frequency between 4 and 6 Hz; 3) rigidity, a raised resistance caused by muscle stiffness during passive joint movement. The resistance exists throughout the range of motion of that joint [190]; 4) postural instability, gradual development of poor balance, which is the primary cause of falls. Other motor symptoms include but are not limited to: speech disorders, respiratory disturbances, hypomimia (masked faces), shuffling gait, festination (shortened and rapid gait), freezing, dystonia (uncontrolled muscle movements), dysphagia (swallowing difficulties), micrographia (handwriting abnormally small) and sialorrhoea (drooling) [101, 166].

QoL is defined as "an individual's perception of their position in life in the context of the culture and value systems in which they live, and in relation to their goals, expectations, standards, and concerns", the World Health Organization [79]. For an incurable disease like PD, all treatment aims to improve the QoL of PD patients. As PD progresses slowly, it is important to help maintain functional abilities through rehabilitation, where QoL plays an important role. Moreover, suitable treatments are based on the correct measurement of PD progression. QoL can provide "important global information for assessing the efficacy of medical interventions" [218]. So investigating QoL can also help understand and improve the treatment of PD.

PD can affect not only people's physical and mental status but also emotional and social functioning [47]. So when treating PD patients, dealing only with physical function is insufficient. The rehabilitation process must consider various aspects of life domains [222]. Non-motor symptoms draw increasing research attention in recent years because they have a tremendous impact on QoL. Mood changes, cognitive decline, sleep disturbance, loss of smell/taste, bowel problems and many other abnormalities are all regarded as non-motor symptoms [174]. From the study's QoL questionnaire with patients, the frequency and severity of non-motor symptoms are the most critical QoL predictor. They even contribute more than motor symptoms to QoL [141]. Non-motor aspects of PD are some of the most troublesome issues that patients perceive [174]. Therefore, studying non-motor symptoms draws more attention to them and has meaningful benefits for PD patients. However, there is a lack of awareness of the importance of non-motor symptoms [25], which social function is a part of. Particularly, social factors play an essential role in QoL [246]. Researchers have compared the QoL of PD patients with the general population and found that one of the areas the disease particularly interferes with is social functioning [194]. Overall, the social domain of QoL and social impact of non-motor symptoms is a significant consequence of PD, but they have never been longitudinally studied before.

Generally, there exists widely used scales to measure PD and its impact on patients. For example, the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) is the gold standard in clinical assessment to quantify PD's overall progression [76]. It is the most commonly used scale in the clinical study [184]. By interview and clinical observations, MDS-UPDRS can comprehensively track the longitudinal course of PD in six different parts. Moreover, the Parkinson's Disease Questionnaire (PDQ-39) or the short version PDQ-8 are widely used tools to evaluate the QoL of PD patients [170]. It focuses on the patients' experience of PD and relies on patients' reports to access eight dimensions of daily living, including social lives, mood and activities. PDQ-39 also is the most frequently used disease-specific health status measure [230]. Nevertheless, compared with the short time in which these scales are conducted, their results are used to estimate patients' situations for an extended period, such as a month or half a year. Moreover, all these scales rely on clinicians' expertise or patients' memories, which are subjective and easily biased [138]. In addition, only a snapshot of disease progression can be recorded, which is short compared with the longitudinally fluctuating disease. Thus, it would be beneficial for both patients and clinicians if continuous and objective monitoring technology could be applied.

2.2 Social withdrawal

Social withdrawal refers to the state in which people avoid the social activities they usually do. It is reflected by solitary behaviours, and it reveals clear social dysfunction [159]. People may avoid social interactions due to social fear and anxiety. Then, they will have less social motivation, which can cause social withdrawal. With the influence of social withdrawal, people will be separated from society, so the disorder symptomatology and deficits in social cognition will be even worse. Personal and social functioning will be seriously influenced, and profound distress and unease will be generated

[40]. It is an essential factor that can cause mental diseases, and it has been identified as one of the main reasons behind mental health-related disability[175]. Social withdrawal is a complex deviation from normal behaviour, which could be influenced not only by disease but also by factors such as age, culture, economic status, availability of transport, mobility [233]. Therefore, studying social withdrawal can provide a novel aspect of understanding and measuring PD, especially non-motor symptoms. The longitudinal observation can generate pathological knowledge of PD, which benefits researchers, clinicians and PD patients' carers. This knowledge can then be used to build necessary therapy so that PD patients can have better QoL and rehabilitation abilities.

Moreover, it is important to mention that social withdrawal is different from loneliness. Humans are social animals. Our primary desire is to form and maintain a minimum quantity of satisfying social relationships [13]. Lack of relationships does not cause loneliness; it is caused by existing relationships that are not satisfying. Loneliness is a type of perceived isolation, which is a highly subjective and emotional feeling [111]. It is identified as a risk factor for morbidity and mortality, which is discussed in [30]. If people suffer social withdrawal for a long time and are not satisfied with the current situation, they will feel lonely. Both physical and mental health will be influenced, and QoL will decrease. Therefore, social withdrawal should be paid attention to in order to maintain social well-being. Although social withdrawal may cause loneliness, and both will reduce QoL, they are still different. [44]. Social withdrawal is a tangible description of the existing absence of social interactions or objective social isolation. In general, social withdrawal is an objective phenomenon. Nevertheless, loneliness is a subjective feeling. Practically, social withdrawal is quantified by a lack of contact with social network members. However, loneliness is determined by subjective social isolation and measured by lack of emotional closeness [157]. Although there is research suggesting that subjective evaluation of social isolation is associated with being objectively isolated [33], this association is not always guaranteed. People may still feel lonely even when they have plenty of connections with others. In this thesis, we rely on smartphones as an objective monitoring tool to observe participants' social activities. So, we limit our scope by objectively measuring reduced social activities, which is social withdrawal.

In addition, everyone has their own social habits and roles. They may have different social patterns and amounts of social activity. There is no universal standard for how much or what kinds of deviation are regarded as social withdrawal. We cannot control all variables to compare social withdrawal purely on certain factors across a population. So, as a complicated phenomenon, it is more likely reached from an individual level.

2.3 Parkinson's disease causes social withdrawal

From the PD perspective, both motor and non-motor symptoms could cause the social withdrawal of PD patients. Moreover, social wellbeing is one of the significant components of QoL, and social withdrawal is a negative side. So by studying social withdrawal, the QoL of PD patients can also be reflected. First are the motor symptoms, including shaking, rigidity, bradykinesia and tremor, which are distinct symptoms of PD. They reduce the mobility of PD patients because it becomes difficult for a person who wants to do exact activities if he/she keeps shaking, has tremors or takes more time when his/her gaits are freezing. Social engagement needs people to go out, join in groups, and do social activities. If basic motor abilities are lost, the possibility of social engagement will reduce, leading to social withdrawal then social isolation. This idea is supported by research that claimed restricted mobility was the defining item for isolation, and both isolation and loneliness are associated with loss of mobility [243] [216] [133]. A study discussing freezing of gait concluded that it has negative social consequences, especially in crowded situations like at the theatre or social events [154].

As for non-motor symptoms, many of them could provoke social withdrawal. In general, they cause difficulties for PD patients to have social interactions. Cognitive impairment is one of the significant non-motor symptoms which can cause social isolation. It was found that PD was associated with impairments of emotional and cognitive social processes, shown by lower empathy and impaired facial emotion recognition [156]. Also, dysexecutive behavioural disorders in the social domain were correlated with the impairment of emotional and cognitive social processes. Because of the cognitive impairment, necessary social functions of PD patients, including emotional recognition and sympathy, are impacted. This problem applies to both facial expressions and vocal tones. Therefore, their abilities of social contact decrease and the potential for social withdrawal increases. The cognitive impairment is even more severe in PD patients with major depression, and they were significantly more impaired cognitively than non-depressed PD patients [212].

In addition, facial masking and dysarthria are reported in PD patients. For facial masking, it causes reduced abilities to modulate and display emotional facial expressions spontaneously [178]. Moreover, there are fewer natural smiles in PD patients than healthy controls [172]. They cannot express their feelings and give reactions accurately through their facial expressions to other people. Misunderstandings may also occur, and they will feel anxious. Over half of PD participants in a study reported facial masking brought the feeling of social distance between themselves and their partners [248]. The more severe facial masking will lead to more significant self-reported social exclusion [82]. In terms of dysarthria, when PD patients want to express themselves by speeches, abnormal rhythm, harsh voice, inappropriate pauses, and prosodic loss could appear [171]. Also, the disabilities in social functions of PD patients cause an extended time in social interactions [191].

These disabilities may also influence the self-perception of PD patients as they feel less competent in communicating with others [145]. This phenomenon was summarised as social anxiety [21]. Even when PD patients with dysarthria try to communicate with others, unfortunately, it is conversation partners who often tend to dominate the conversation [189]. They will feel less represented and respected. PD patients may fear situations such as interactions with other people because they are afraid of being judged and evaluated. Even worse, social phobia was detected in some PD patients [81]. Public appearances could be a shame for them, so they decrease social exposure, which then leads to social withdrawal [160].

All these deficits could make PD patients not motivated to have social activities. So stigma and apathy were discovered in PD patients. They have to face felt stigmas, such as shame, embarrassment, and disgrace, and enacted stigma when encountering others, such as staring, questioning, and avoiding [134]. These negative responses all exacerbate the disguising of symptoms of PD patients. It includes reduced interest and participation in normal purposeful behaviour, lack of initiative, and problems in initiating or sustaining an activity to completion, which is termed apathy [173]. The PD patients have no sense of security in social activities, so they may reduce the frequency of social interactions and avoid communicating with others [86]. So they may later withdraw from the public into a closed world when their symptoms can no longer be hidden. Overall, various kinds of PD symptoms, especially non-motor ones, can cause the social withdrawal of PD patients.

2.4 The feasibility of smartphone sensing

Observation is the foundation of social behaviour research [39]. It is the prerequisite to any analysis or investigation. Observation methods are diverse. Different observation methods can be implemented according to research projects and objectives. Generally, there are two types of observation: active and passive. The active observation approach needs active input from researchers or participants in the data collection process, which may add extra burden and influence participants' reactions [106]. Compared with active observation, passive observation does not directly affect participants' behaviour, which is advantageous in particular studies. A comparison of active and passive observation is beyond the scope of this thesis. So, we limit our scope to passive observation.

Passive observations can benefit from the unobtrusiveness of the methodology. It reduces influence on the experimental targets to alleviate their reactions towards the observation results. There are no specific tasks for participants to complete, which relieves them of any burden. Participants can continue their behaviours without paying attention to the observation process. So, the observation can be ubiquitous, and participants' ordinary lives can be monitored in the wild. No special participation in the observation process can also prevent bias from subjective consciousness. Unobtrusive sensors can help to achieve passive observation and release the Hawthorne effect [201]. In addition, the unobtrusiveness and ubiquitousness of passive observation makes longitudinal studies feasible. The cost of active sensing is expensive in long-term research. For example, participants may have to travel to specific places for every measurement, and the number of qualified examiners is limited [49]. In this case, passive sensing is more appropriate because the data can be collected automatically without extra operations. It is also especially advantageous for observation of people with mental diseases, such as dementia and schizophrenia [43], since no additional cognitive burden is added.

Technologies like Radio-frequency Identification (RFID) and wearable sensors achieve the goal of passive sensing [14, 69]. Face-to-face social interaction can be explored by RFID [9]. However, RFID is an extra device that participants have to carry. They are still external devices which differ from typical consumer products like smartphones.

If participants are asked to wear them, they may still feel monitored and behave differently in their daily lives. So results from studies using external devices may differ from natural settings. Compared with existing equipment, designing and manufacturing an external device costs extra time and money. Typical wearable sensors include smartwatches, wristbands and activity trackers. Health data, such as blood pressure, heartbeat, steps, time standing, heights of climbing, oxyhaemoglobin value and sleep quality, can be obtained through wearable sensors. Because of their rich functions, wearable sensors are increasingly popular among consumers and researchers, but their penetration rate is still much lower than smartphones. In 2020, only 41% of UK households owned wearable sensors, while 97% of them used smartphones [215]. Moreover, wearable sensors are also noticeable, especially for people who do not usually wear them [192].

Based on penetration rate and compared with other passive sensing devices, smartphones are less intrusive for collecting social behaviour data. Unlike wearable sensors or RFID, which need pairing with other devices, no extra equipment is required in smartphone sensing. Smartphones are not only owned by the majority of people, they also play an important role in people's lives. They have also become the social hub for personal communication. In 2020, UK adults spent an average of 2 hours and 34 minutes online through their smartphones [68]. It has potential in many areas, such as sociological, psychological and medical research. With miscellaneous functions and sensors carried by the smartphone, various types of data can be captured and inferred. The communication mediated by smartphones, such as calls, messages [195] and social media [74], can be captured to analyse the social activities of the users. Furthermore, the data from the sensors can also reflect the surrounding environment. Information from the microphones [239], Global Positioning System (GPS) and Bluetooth [250] can infer conversation and social factors in the real world. Such data has been used in the research of general human behaviour [239], loneliness [43], personality effects [35], work efficiency [59], healthcare [43, 55, 119], diseases such as schizophrenia [28] and transportation and behaviour measurement [120, 109]. Its feasibility and validity in behaviour research have been confirmed [18]. Smartphone social sensing has a promising future because both contextual and behavioural data can be collected so that behavioural changes can be monitored from a more comprehensive perspective [11, 202].

To the best of our knowledge, PD-caused social withdrawal has never been studied using smartphone social sensing. The analysis of social behaviour is essential to PD, as social withdrawal can reflect disease progression and QoL of patients. Therefore, the smartphone can continuously and unobtrusively monitor PD patients' social lives so their status of social withdrawal can be thoroughly understood, which can then act as a reference for treatment and personalised care. We also conducted a systematic review on passive smartphone social sensing in Chapter 3, which completely summarised the details of this technology. It can be regarded as part of the background and related work.

2.5 Related work: Digital phenotying in Parkinson's

Digital phenotyping is 'moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices' [164]. As mentioned before, smartphone is feasible for monitoring users' social behaviours, and it is a particular type of digital phenotyping, in which the digital device is the smartphone. Digital phenotyping has been applied in PD research for various purposes, such as differentiating PD patients and healthy controls [118] and monitoring symptom progression [8]. There are six identified clinical problems (i.e. improving diagnosis, monitoring response to therapy and motor-symptom fluctuations, monitoring non-motor symptoms and progression, improving medical treatment, enhancing surgical treatment and improving rehabilitation interventions) that need technologies relevant to the diagnosis and clinical management of patients with PD [61]. Given the tools of monitoring and aiming to understand PD in this work, we focus on smartphone-based continuous monitoring of motor and non-motor symptoms of PD. Studies that classify healthy controls and PD patients or use smartphones as a questionnaire delivery method are excluded.

The previous study also summarised four categories of technology-based PD monitoring according to the activeness of the measurement environment [94]. They are 1) active and controlled: active sensing approaches (i.e. including human contact) in controlled environments where participants must accomplish designated tasks in the laboratory or in a home-like setting; 2) active and in-the-wild: active sensing approaches in the wild, in which participants are required to undertake designated tasks as part of their daily routines; 3) passive and controlled: passive sensing approaches (i.e.

without human contact) where participants are being monitored in controlled environments or home-like settings; 4) passive and in-the-wild: passive sensing approaches in which participants continue with their daily routines without interruptions. Although controlled studies have achieved promising results, their results are based on the elimination of broader influences, and results' generalisability were degraded in real-life situations [205]. Circumstances in the wild are much more complex than controlled environments because of the extensive variables introduced. If studies always remain in the laboratory, it will impede the development of sensing systems with practical value for border application. Thus, only smartphone-sensing studies that were conducted in the wild are included.

There have been several systematic and non-systematic reviews on smartphone-based assessment and monitoring of Parkinson's [205, 127, 129, 176, 2], which are general references for the reader. We also conducted the keyword search with 'smartphone Parkinson's' to retrieve studies using smartphones to study PD knowledge. Collectively, 12 pieces of literature are included in this review. Smartphone monitoring also has the potential to benefit from four novelties of this sensing method: unobtrusiveness, longitude, personalisation and non-motor-symptom applicability [94]. These features also led us to the vision of building a clinical monitoring method valid in real-world settings.

2.5.1 Unobtrusiveness

Unobtrusiveness is the opposite of burdens on participants and examiners. It does not require participants' direct input during the experiment. Compared with external sensors, the smartphone has original unobtrusiveness because it is popular among the general population. Studies can benefit from its ubiquity and familiarity. Typically, studies employ specific motor tasks, such as sustained phonation, rest tremor, postural tremor, finger tapping, balance and gait, via smartphones as symptom assessments [128]. These tasks usually ask participants to leave their regular activities to complete. Therefore, these burdens impact the compliance rate of application usage. The evidence shows that '26% of apps are used only once, and 74% of apps are not used more than 10 times' [61].

Zhan et al. [254] recruited 121 patients and 105 controls to study the dopaminergic medication response. Participants were asked to complete five smartphone-based activities (each likely requiring five minutes) using the HopkinsPD app twice daily before and after medication. They were 71% accurate in detecting motor changes in response to medication. Nevertheless, 7,563 sessions were recorded over a six-month period, equating to an average of 33 sessions per participant. Only 41 patients with Parkinson's participated in more than 20 evaluation sessions. Additionally, a daily set of 17 smartphone-based tasks was requested of 35 patients to complete as part of cloudUPDRS system development. However, the compliance rate dropped sharply after the first week, and only 12 patients (34%) completed the research after 12 weeks [210].

Although other studies implemented a reduced number of tests or shorter test sessions for participants to perform, high compliance rate of participants is not guaranteed. For a study aiming to assess the feasibility of incorporating the smartphone research application into a PD clinical trial, a set of seven active tasks (finger tapping, standing still for 30 s, walking for 30 s, saying 'ah' for 10 s, assessments of rest and postural tremor and a brief spatial memory test) were implemented to assess PD symptoms [165]. However, only 61% of participants completed at least one smartphone task at three months, 54% at six months, and 35% at 12 months. Moreover, a five-minute set of five activities four times daily for a month was requested of 20 participants to support expert monitoring of PD in Aroral et al.'s study [8]. The results showed only a 68.9% compliance rate, and an average of 2.7 tests per day was achieved.

In contrast, unobtrusive studies do not usually suffer reduced participation in examining sessions because no repetitive test on smartphones is involved. With seven participants of earlier Parkinson's, Julio [94] developed an adaptive system for tracking self-reported variations in pain, gait and fatigue on a weekly basis. His strategy outperformed chance in tracking fluctuations with Cohen's kappa values ranging between 0.19 and 0.53, indicating that each participant's subset of predictions was unique. None of his participants dropped out due to repetitive input on the smartphone because all monitoring was unobtrusive.

2.5.2 Longitudinal

Accurate data on the long-term evolution of PD symptoms and their short-term variations are critical for ensuring appropriate treatment for PD patients and accurately measuring the results of clinical studies [20]. Therefore, studies focusing on PD progression had longitudinal settings to observe changes of PD patients over time. However, due to extended experiment time, active longitudinal studies still suffer from a reduced compliance rate.

In order to determine the feasibility, reliability and validity of smartphone-based digital biomarkers of PD in a clinical trial environment, 44 individuals with PD were requested to complete six daily motor-activity assessments for six months [128]. All PD participants completed 61% of all possible test sessions throughout the six-month study period. 64% of individuals with Parkinson's disease completed all active tests at least once every other day, and 90% completed them at least once every four days. However, in the last week of the six months, participants with PD completed 43% of tests on average each week. In the mPower study [22], which had a large number of participants, participants were required to complete four smartphone-based activities three times a day in brief sessions. After six months, just 150 of the 1,087 persons who enrolled and self-identified as people living with Parkinson's utilised the app on at least five separate days. Moreover, the monitoring application was barely used by any participants after 100 days.

Potentially, there could be a link between the complexity of tasks and the drop-out rate of participants. Generally, participants are less likely to remain interested over time with increasing complexity of tasks. Thus, if participants are involved in actively monitoring evaluations during the longitudinal experiment, it is critical to maintain a moderate burden for them[94]. In other words, if the measurement is unobtrusive (i.e. no active input from participants) an extended length of time observation and a high adherence rate can be achieved. Liddle et al. [124] defined Lifespace as a multidimensional geographic space where people live and do activities, which reflects mobility, health, and well-being. Through a visual comparison, a study discovered 'a trend of decreasing Lifespace with increasing severity of reported symptoms as measured by the initial partial UPDRS score' [124]. Although it is a proof-of-concept study, it revealed that Lifespace could be a sensible objective outcome measure incorporated in a person's local environment for monitoring PD patients' lived community access and engagement. Vega [94] also achieved longitudinal observation of seven PD participants over a year. He only asked participants to complete a simple daily diary as ground truth. To answer the diary, participants just needed to fill the tiny dot of the severity of that symptom [234]. Except for that, all smartphone data was collected unobtrusively, and no functional test was required. The compliance of diaries only declined by 2.15% over 299 days.

High cost and burden have caused the absence of naturalistic and longitudinal studies based on active monitoring. Participants are unlikely to put up with invasive monitoring devices or time-consuming evaluation activities for an extended time. Nonetheless, because of PD's complexity, longitudinal monitoring is necessary to provide a more precise picture.

2.5.3 Personalisation

Precision medicine (or customised medicine) is defined as 'health treatment that is tailored to an individual based on their genes, lifestyle, and environment'. As a complex and diverse neurodegenerative condition with a wide variety of symptoms (motor and non-motor) and side effects from treatment, PD is likely to benefit from a precision medicine approach [186]. From reviewed studies, personalisation depends on the method researchers apply.

In the PDLens study [256], 81 PD participants completed gait, balance and voice tests to detect drug effectiveness during daily life over six months. A deep neural network was created to establish both person-centred and person-independent models. Their system was capable of predicting the conditions before and after drug intake with 70.0% accuracy in person-independent results. However, this performance was lower than that of the person-centred model. Person-independent medication intake detection is significantly more difficult, as pharmacological efficacy can vary significantly between participants due to illness progression and medical prescriptions. Two hundred forty-seven participants with PD and more than 100,000 gait-cycle samples were included in the PDMove study [260] to explore medicine adherence of PD patients. To achieve personalised models, they adopted transfer learning (i.e. a technique in which a model trained from one task is repurposed for another related task), which allows a deep learning model to work with a small number of data. Personalised adaptation was also applied in the cloudUPDRS study [210]. After a one-week calibration period, they trimmed down 17 active smartphone-based activities to a customised subset that accurately predicted individual UPDRS scores. Their objective was to shorten the time of their evaluation sessions in order to boost compliance. They justified their experiment by emphasising PD's idiosyncratic nature, since each patient often develops a collection of symptoms that dominates their UPDRS scores. They address personalisation by establishing a machine learning approach for personalising assessments which pick test sections that most closely fit individual symptom profiles. Thus, the one with the most excellent inferential power was selected, so it could accurately determine the patient's overall score.

By identifying regions of dysfunction and their link to therapy, the smartphone sensing approach is utilised to deliver personalised feedback to individual patients and perhaps stratify factors that 'predict' a patient's reaction to various treatment paradigms [61].

2.5.4 Applicability to non-motor symptoms

Non-motor symptoms, however, are frequently overlooked [25]. According to the QoL questionnaire, the most significant factors affecting QoL are the frequency and intensity of non-motor symptoms. They are even more significant than motor symptoms [141]. Non-motor deficiencies frequently influence patient priorities and causes of disability (e.g., depression, anxiety, fatigue, orthostatic hypotension, sleep disturbance). There is an urgent need to develop unobtrusive technologies for monitoring non-motor endpoints in the home and community.

However, almost all studies we found focused on tracking motor symptoms, such as tremor, freezing of gait, hand movement, postural stability and voice disturbances. The ground truth they relied on was usually partial or total UPDRS scores. There were only a few studies that monitored non-motor symptoms. Memory tests were conducted in Prince et al.'s study as part of the mPower project [179]. Participants' short- and long-term behaviours were analysed, and non-significant impairment but 'a larger degree of longitudinal performance variability' were found between PD participants and healthy controls. Pain and fatigue fluctuations were monitored in Julio's study [94]. He utilised unobtrusive mobility data and activity-recognition metrics from smartphones to track these symptoms.

2.6 Conclusion

This chapter introduced PD first. Then, social withdrawal and how PD causes social withdrawal were explained. Social sensing is the basis of analysing people's social lives, so we also discussed methods of social sensing. As smartphones are popular among the general population, no extra cost or equipment is required to collect data. We highlight the applicability of unobtrusive, longitudinal, personalised and non-motor symptoms to the smartphone social-sensing method. As PD is complex and the symptoms fluctuate, PD monitoring should be longitudinal and continuous. Digital phenotyping has been confirmed to be reasonable when keeping track of the motor symptoms of PD. However, as an overall indicator and essential part of QoL, social withdrawal has not been studied in a longitudinal manner. Therefore, we propose the method of using a smartphone to monitor the social withdrawal of PD patients. The detailed methodology and experiments are described in the following chapters.

Chapter 3

Passive Social Sensing with Smartphone

The background chapter discussed that Parkinson's disease could cause social withdrawal, which practically implies reduced social interactions. Therefore, the actual variable we are measuring is social interaction changes. In addition, to minimise the awareness of monitoring, the measurement has to be passive and unobtrusive. As the social hub, smartphones compose a considerable amount of an individual's social interactions. Combined with other sensors, they are also capable of inferring social interactions that occur outside the smartphone. In addition, using a popular device for monitoring is less intrusive than introducing novel ones. After choosing the smartphone as our primary monitoring tool, the first question became how to measure social interactions via smartphones. However, to the best of our knowledge, no literature collected all information together and gave an overview of how smartphone passive social sensing was conducted. To answer these questions reasonably and to understand this concept thoroughly, a systematic review was initiated to fill the gap. It discovered a paradigm of digital phenotyping experiment, including data collection, ground truth retrieval, and relation interpretation. Critical elements of this technology, such as sensors utilised, strategies applied, features extracted, and data analysed, are summarised in the review. In the following chapters, we adopt the smartphone passive social sensing to PD patients in the wild.

The content of this chapter is adapted from *Heng Zhang, Ahmed Ibrahim, Bijan Parsia, Ellen Poliakoff, and Simon Harper. 'Passive socialsensing with smartphone: A systematic review'*. It's currently under review.

Author's contributions

Heng Zhang designed and conducted the systematic review, summarised the results, discussed the findings and wrote the manuscript. Ahmed Ibrahim participated in cross-check of the quality of included literature. Ellen Polikoff, Bijan Parsia and Simon Harper provided the guidance of the review and gave suggestions on the writing of the manuscript.

Abstract

Background: Smartphones are widely used and have become hubs of personal communication. Combined with multiple sensors, they are capable of capturing social interaction on and outside smartphones passively. So it is feasible to apply smartphones as novel research tools to conduct social-related studies.

Objective: To review the published empirical English literature of passive social sensing with smartphones. To explore which domains this technology apply, and its sensing strategies. Their performance in practical studies, benefits and challenges will also be discussed.

Methods: Following the PRISMA guidelines, we constructed a search string considering variations of the term smartphone. The search was conducted in ACM Digital Library, IEEE Xplore, PubMed, ScienceDirect and Web of Science. Snowballing method was also applied following Wohlin guidelines after the initial search. Papers were included if they were empirical, only used sensors on smartphones to collect data, involve measuring social interaction or social activity level, required minimum user interaction and described as passive.

Results: The search produced 2741 results, of which 47 eligible articles were identified, two more articles were added from the snowballing. Participants of included studies ranged from 5 to 11000 and described experiments length range from two weeks to two years. The aim of all studies was to understand human social behaviours, and some of them correlated social interaction levels with other topics such as diseases, wellbeing and personality. College students were the most common participants. The most popular operating system was Android. Calls, messages and Bluetooth data were the three most frequently used sensors. Various data analysis methods, including simple as correlation, complex as machine learning, have been utilized. All studies achieved their aims, And the most common approach to represent findings are correlations. Benefits such as ubiquitousness, unobtrusiveness, personalisable, and continuity and challenges such as privacy, accuracy and methodology were reported and summarized from

reviewed studies.

Discussion: Other than questionnaires, passive smartphone social sensing gives researchers another perspective to understand participants' social lives. Reviewed studies confirmed the feasibility and validity of this technology. However, they still suffered from privacy concerns of participants, relative sample sizes, the significance of experiments is not generalisable and data integrity due to technical faults. The recommendation is to conduct more research on making reasonable sensor frequency choices, standardising smartphone features, building personalized models and implementing state-of-art technologies.

Conclusion: The smartphone social sensing technology provides innovative opportunities to measure human social behaviour objectively. It has a promising future in the field of sociological, psychological and medical researches. With concerns including privacy, accuracy and methodology, its evolvement needs to be addressed by interdisciplinary collaborations between technology experts, computer scientists and professionals in all related fields.

3.1 Introduction

Human beings are social animals. Social activities have taken an important part of our daily lives, and it has a significant influence on people's mental and physical health. Poor quality of social relationships is a major risk for depression [225]. There is evidence showing that socially isolated people have a higher mortality rate than the general population [216]. On the contrary, positive social behaviour is beneficial to both individual and society. For example, efficient social interaction can save time and increase the productivity of workers in certain industries [7].

Consequently, social behaviour research has begun in the last century [91]. Typically, experiments on social behaviour consist of three parts: 1) set particular contexts: for example, arrange a scenario making people nervous 2)observe participants behaviour: record their voice, face emotion and reaction of social interactions under this scenario, and 3) retrieve feedbacks from them: use scales, questionnaires or interviews to measure their mood or satisfaction of this social behaviour [235]. Methods of experiments include interviews, questionnaires, voice or video recordings and expert observations [167]. However, these approaches come with their natural defects and limitations.

Questionnaires could introduce errors by recall bias; participants may behave differently in aware or controlled settings; the cost of these techniques could restrict the scale of the experiments [80].

These active observations may disturb participants in a variety of ways, but passive sensing has opportunities to reduce disturbance and burden fundamentally. It does not require intensive inputs from participants, so it is unobtrusive compared with traditional measures and all data are captured in situ [207]. Passive sensing refers to different kinds of method to collect data from participants by minimum or no direct interaction with any object or person. Participants do not have to do specific tasks or answer particular questions during or after the experiment. Not only the burden on participants' awareness, cognition and memory are released, but also the recall bias and Hawthorne effect are alleviated if unobtrusive sensors are applied [201]. Moreover, the longitudinal study becomes more feasible by combined unobtrusiveness and ubiquitousness of passive sensing. Data can be collected continuously without the presence of observers or laboratory environment but in the wild. It is beneficial for sensitive and stigmatised social experiments in mental health, such as studies in dementia and schizophrenia [43].

In some studies, dedicated research quality technologies of passive sensing have been proposed and examined in obtaining social activity data. A predecessor is in the auditory domain, the electronically activated recorder was invented to sample the surrounding sound of participants to infer social activities and conversations [142]. Another example is RFID devices. They were designed as badges with beacons to explore face-to-face social interaction between individuals [9]. They could achieve high accuracy and detailed mapping of every social interaction [59]. In the RFID studies, these specially designed RFID badges are worn at the front of participants' bodies so that the signals can be easily captured. Particular scanners are also working continuously to capture the RFID signals and infer if participants have social interactions in a specific room. All these procedures can maximise the chance that RFID signals can be accurately captured. As for the Bluetooth on smartphones, it can only scan surrounding signals at a certain frequency, so the signals in between could be lost. Moreover, some users may also not turn their Bluetooth on, so the smartphone will lose detection of these people. Furthermore, smartphones are usually kept in pockets, so the Bluetooth signal might interfere, which then leads to signal loss. Overall, dedicated RFID badges are more reliable and accurate than Bluetooth on smartphones when detecting surrounding people in a specific area. However, in contrast to using existing devices, RFID could raise the cost of experiments. Comparing with usually carried gadgets such as smartphone or smartwatches, they could be more obtrusiveness because people may notice these badges. Recently, wearable sensors(e.g. smartwatches, wristband activity tracker) are becoming prevalent for extracting behavioural cues. They have dedicated sensors to monitor more targeted variables such as heart rate and stages of sleep. But they often paired and existed with smartphones all the time, and their penetration rate is much lower than smartphones. In 2020, In 2020, at least one person in 97% of UK households own a smartphone, only less than a half, 41% of them have wearable [214]. Moreover, these technologies also could make participants aware and may bring extra reaction towards sensor wearing people. Comparing with smartphones, they have a potential ecological threat for people who do not usually wear them [192]. Indeed, all kinds of sensing methods have their own benefits and limitations. Studies can choose the applicable one or combination of tools according to the aims of their studies. As an off-shelf device, smartphone has the highest ownership among all of them. For majority of people, they do not need extra equipment to participate these studies. It is convenient for experiments in-the-wild. We limit our scope to smartphones.

Nowadays, smartphones have become the hub of personal communication and computing. In 2020, 87% of UK adults owned smartphones [161]. Even in people aged over 55, ownership rates of smartphones rise significantly from 4% to 70% from 2008 to 2020 [162]. It is reported that UK adults spent 2 hours and 34 minutes on average day online on their smartphones in 2020 [68]. Social interaction happened on the smartphone, such as calls, messages, emails, and social media activity can be captured on it naturally. So social sensing by smartphone could be less intrusive than any other devices.

Additionally, off-the-shelf smartphones are embedded with multiple and power sensors. These sensors empower smartphone as an efficient tool to capture not only social interaction mediated by smartphones but also the surrounding social context. For example, raw data from sensors such as microphones, Global Positioning System (GPS), accelerometers can be gathered and interpreted as conversation engagement, mobility patterns, number of encounters to inference social interaction happened outside of smartphones. These data can then be analysed to assess related topics such as depression, loneliness [43] and work efficiency [7]. Moreover, combined with the capability of storage, process and off-load data, smartphones can be set up easily to accomplish these tasks.

Although several surveys or reviews have given big pictures of how smartphones passive sensing can be utilised for various fields including healthcare (e. g. [43] [55]), transportation and behaviour measurement (e.g. [120] [109]), they did not explicitly discuss how this technology were applied for social interaction measurement. To our knowledge, passive social sensing by smartphones is not systematically reviewed. The goal of this article is to summarise the existing literature and address this gap.

3.2 Objectivess

The main objective of the systematic review was to explore how smartphones were applied for passive social sensing. Detailed research questions are:

- To which domains and populations have smartphone passive social sensing been applied?
- Which kinds of sensors and data collection methods have been used for smartphone passive social sensing?
- How were sensor data analysed after data gathering?
- What is the accuracy or other performance indexes of these passive social sensing methods?
- What potential problems did smartphone social sensing had and were there any solutions?

3.3 Methods

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [149] guidelines to perform the systematic review. After the final decision of included papers was made, we followed Wohlin guidelines [247] for snowballing in systematic reviews using those papers as the starter set.

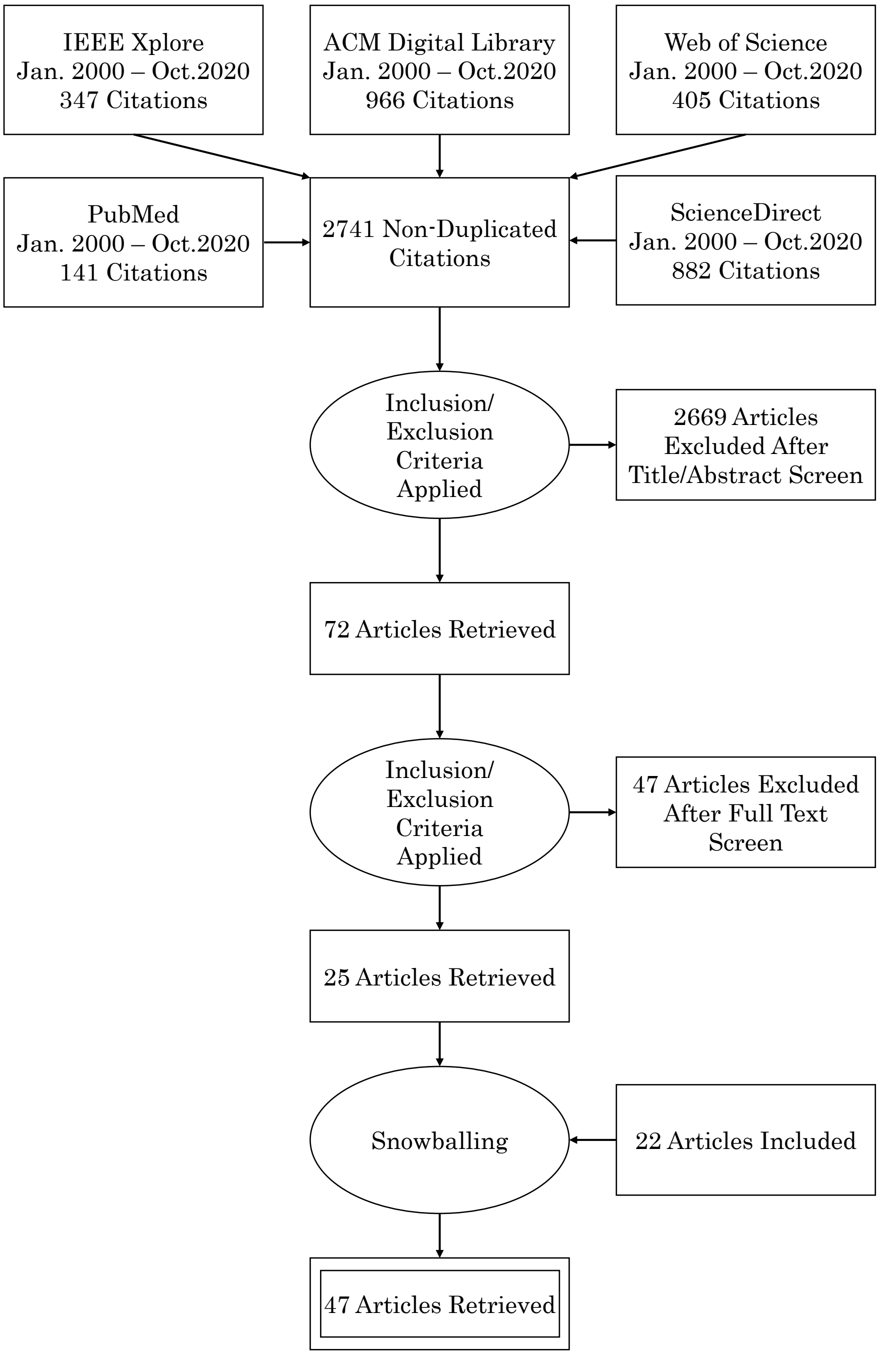


Figure 3.1: Diagram of the review process.

3.3.1 Types of studies

Studies were included if they were:

1) empirical studies, involve experiments on humans;

2) use sensors embedded on the smartphone only;

3) the aim or indirect aim of the sensing is to detect if users of the phone engage in social interaction or their overall social connectedness level (Social interaction means 'the process of reciprocal influence exercised by individuals over one another during social encounters'. It refers to face-to-face encounters in which people are physically present with one another and technologically mediated like calling or messaging [130].);

4) involve data collection on smartphone;

5) require minimum user interaction on the smartphone, described as passive.

Studies were excluded if they were:

1) crowed sensing, because it combines all sensors' data from a large number of smartphone users but not an individual;

2) using other sensors such as wearable devices paired with smartphones, special fixed sensors at home because they did not use smartphone sensors;

3) require participants to put their phones in specific positions such as body, clothing because they are intrusive and only use smartphones as accessible sensors.

We defined smartphones as mobile phones running an operating system including but not limited to Windows Mobile, Symbian, Android, iOS, which third-party applications can be installed for data collection purpose. Passive was regarded as data collected without user input, except the data collected for building ground truth such as the target for correlation and labels for machine learning.

Although there are other external sensors and scenarios can be applied on smartphones studies, the search scope is strictly limited to smartphones because they are different from normal smartphone usage settings in-the-wild. It could threaten the generalisation and passiveness of results. In addition, papers focus on specific technology only such as analysing raw audio for vocal inference or constructing proximity networks from Bluetooth signals were not included. Because their aims are exploring these technologies rather than understanding human social interaction behaviour.

We included English-language peer-reviewed journal papers and conference proceedings published from January 2000 to October 2020. We choose to start from 2000 is because it was the first time Bluetooth is embedded in smartphones, which enables smartphones as inference for surrounding social interactions.

3.3.2 Search strategy

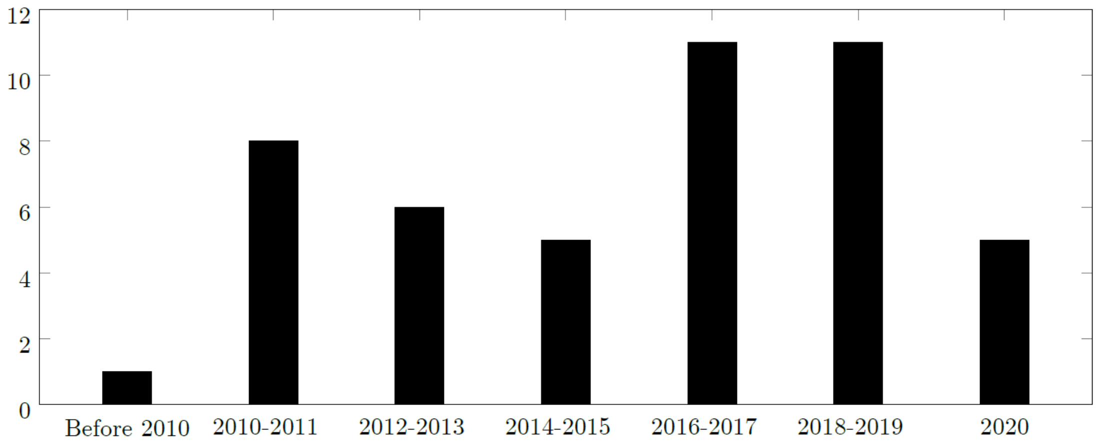
We conducted two searches in computer science and electronics domain-specific databases ACM and IEEE, one search in health domain PubMed, and two searches in the cross-domain database Web of Science and ScienceDirect. We build the search string based on our inclusion and exclusion criteria with special consideration of variations of the term smartphone.We've carefully considered possible related research topics of passive social sensing but realised the fact it is the first of its kind and no literature can tell how many terms are necessary to cover the whole area. So we can not pre-defined all possible fields passive smartphone social sensing was applied and added them as keywords in search terms. Also, one of the objectives of the review is to explore which domains and populations have smartphone passive social sensing been applied. Therefore we choose the reasonable way, use only social and sensing as two keywords. We did not include the term 'passive' in the search string because according to the pre-search, a number of qualified studies can not be retrieved by search engines, which probably because they did not have the word 'passive' in their title or abstract. So we decided to execute the search string without it but select results manually.

The search string was: (smartphone OR cellphone OR cell-phone OR 'cell phone' OR 'cellular phone' OR 'mobile phone' OR 'mobile telephone' OR iPhone OR iOS OR Android OR Symbian OR 'Windows phone') AND social AND sensing NOT (crowd OR community)

3.4 Results

We included a total of 47 publications and the summary of them are shown in table 3.1. They were selected from 2741 non-duplicate results from five databases mentioned above after title/abstract screened and full-text reading. These results were examined and discussed by the first and the second author according to the inclusion and exclusion criteria. The whole process was shown in figure 3.1. In the snowballing procedure, the total number of citations of a particular publication exceeds 2,000, but most of them did not meet our inclusion criteria. To improve efficiency, we execute

Figure 3.2: Studies by year of publication



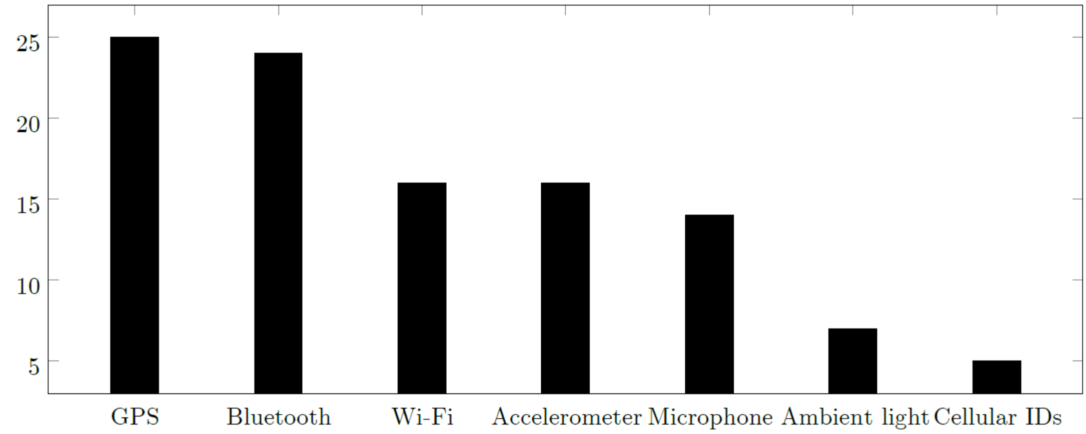
our searching string again on those citations to narrow results. In addition, studies doing secondary analysis on public social datasets not collected by authors but other researchers were also considered. These public datasets include Lausanne data collection campaign dataset [112], reality mining dataset [57] and StudentLife dataset [239].

The majority of these citations were discarded because of irrelevance (e.g. studies in robotics), theory only (e.g. conceptual papers of privacy and algorithm in social sensing), intrusiveness (e.g. they asked participants to label data before the experiment started) and involve sensors other than smartphones (e.g. fixed locating sensors or wearable sensors). Although some studies used smartphones passively, they were still excluded because their targets are mobility or context of users but not social interaction.

Twenty-two studies (47%) were conducted in the United States; 15 studies (32%) were in Germany, other studies were performed in Switzerland, Italy, the United Kingdom, Denmark and China. Moreover, 27 papers (57%) were published after 2015. Details are shown in figure 3.2.

As for the intentions of these studies, 21 (45%) are understanding human social behaviours, which includes proximity detection, relationships evolution, etc.. Personality is also a favourite topic, which catches 13 (28%) of reviewed studies. Other studies correlate social interaction levels with well-being, involving depression, anxiety, stress, obesity, sleep problem and mood. Particularly, four studies applied smartphone social sensing for disease research, including schizophrenia and bipolar disorder [66] [240]

Figure 3.3: Applied physical sensor of reviewed studies.



[29] [238]. Active measures for participants were involved in reviewed studies. For example, Wahle et al introduced an intervention for participants to alleviate maladaptive thinking[236].

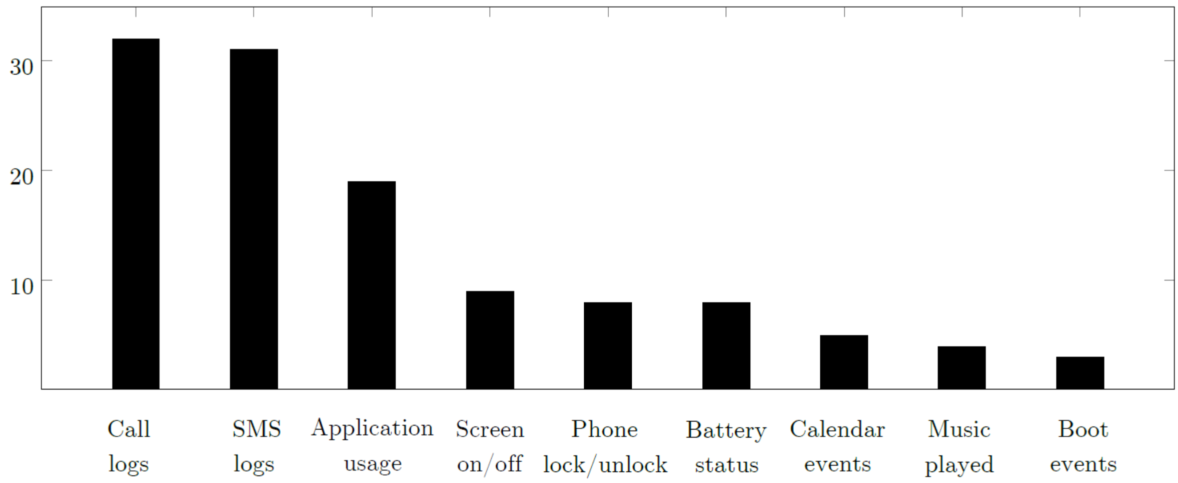
46 (98 %) selected studies reported the number of participants. The number of analysed participants of ranges from 5 to 11,000, the mean is 380 with the standard deviation of 1613 and median of 54. Length of experiments were explicitly described in 44 (94%) studies. 35 (74%) studies had a fixed study duration, and the study length ranges from 2 weeks to 2 years. The average study time in days is 184 with a standard deviation of 207 and median of 70. As mentioned above, most experiments were conducted in developed countries, where smartphones are prevalent.

Seven studies did investigations on datasets collected by others. Lausanne data collection campaign dataset [112] is the most interpreted one. It was investigated in five studies [54] [53] [83] [35] [36]. Reality mining dataset is originally reported by Eagle and Pentland [57] and analysed in two studies [64] [253]. The result of two studies [228] [237] came from the StudentLife dataset [239]. Two studies [29] [28] shared the same data.

3.4.1 The paradigm

From all reviewed studies, a paradigm of passive smartphone social sensing can be summarized. Typically, an application will be installed on participants' smartphones,

Figure 3.4: Applied on-device analytics of reviewed studies.



and it will collect designated sensor data throughout the experiment period. The gathered data is transmitted to a remote server or stored locally on smartphones. Simultaneously, ground truth such as clinical/psychological scales or self-designed questions will be conducted at the beginning, the end or during the experiment at a particular frequency. After the data period, raw smartphone data will be processed, and higher-level features will be constructed from them. Then, various analysis methods will be applied to discover the relationships between ground truth and smartphone data. The overview of the paradigm can be seen in figure 3.5.

All reviewed studies implemented an observation approach to investigate their research questions. For health-related studies, none of them had a detailed hypothesis that certain features from smartphones have particular relationships with their research targets. They implemented smartphone passive social sensing technology and analysed all collected data with the ground truth afterwards. The results showed that all studies had found all or some of smartphone features can indicate or associate with research objectives.

3.4.2 Participants

Participants are foundations of the empirical researches. Although only one study reported that sample size was determined based on an a priori power analysis [195], all selected studies report the population of their experiments. College students and staff

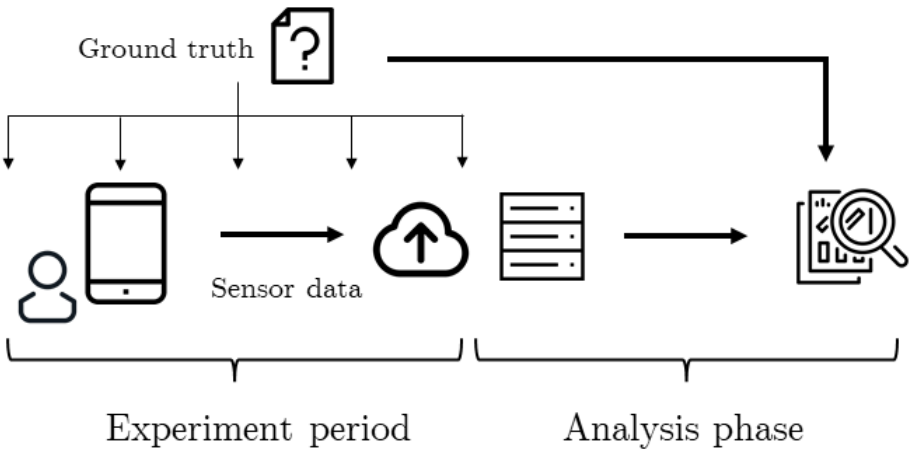


Figure 3.5: The paradigm of passive social sensing with smartphone.

including undergraduate, masters, PhDs and researchers are the most prevalent one among all papers, which takes 23 (48%) of them. Ten of which even either work in the same laboratory/university or live in the same dormitory building [57] [64] [163] [253] [143] [216] [114] [228] [137] [135]. Two studies recruited young adults ranges from 18-21 in the surrounding area of the university [202] [11]. Relationships of participants in six studies are colleagues, friends or family members [54] [53] [83] [35] [36] [3]. Participants in two studies are typical families consists of parents and children. In Centellegher et al's study, each family lived in different regions [32]. However, in Moturu el al's study, participants resided in the neighbourhood community [155]. For other studies, potential social relationships of participants were not clarified. Moreover, disease-related papers all had special criteria for recruiting participants. For example, Buck et al's studies [28] [29] only include candidates who had a diagnosis of schizophrenia.

Only twelve studies (23%) explained the provider of smartphones which participants used for experiments. Seven studies (15%) gave smartphones to their participants. Two studies helped participants migrate to new phones [238] [12]. However, participants in [239] treated given phones as secondary ones. Five studies (11%) studies install sensing applications on users' own smartphones. Especially, Servia et al asked participants to download the application themselves to take part in the study [199] .

For longitudinal experiments with human involvement, there is always a risk of participants withdrawal. Only three of selected studies reported details of participants dropping-out. One study reported 50.8% of participants uninstalled the data collection app within the first two weeks with only one-fifth of participants left for four weeks [236]. Buck et al [29] reported 8 participants (13% of total) withdrawal, and 14 (29% of total) in Wangs et al's study [238]. Possible reasons are not qualified, no longer interested and felt required too much effort.

3.4.3 Sensors

A variety of sources on smartphones are being applied to capture different types of data, including both physical sensors and on-device analytics. The most utilised physical sensor is Global Positioning System (GPS; 25 studies), and Bluetooth, which was applied in 24 selected studies. Other prevalent sensors are Wi-Fi (16 studies), accelerometer (17 studies) microphone (14 studies), ambient light (7 studies) and cellular IDs (5 studies). Other data are from on-device analytic, which includes call logs (32 studies), Short Message Service (SMS) logs(30 studies), application usage (19 studies), screen on/off (8 studies), phone lock/unlock (8 studies), battery status (8 studies), calendar event (5 studies), music played (4 studies), and boot event (3 studies). GPS is generally employed for positions in most scenarios, but individually, locations are obtained in different mechanisms. Two studies referenced visited places of participants by detecting if their smartphones connected with certain Wi-Fis [216] [137]. One study collected Wi-Fi, GPS and cellular ID for locations [199]. Two studies combined Wi-Fi and GPS to observe location behaviours [12] [237] . Besides, three studies adopted cell tower IDs to interpret participants' rough neighbourhoods [57] [64] [253]. Details of sensors usages of reviewed studies are shown in figure 3.3 and figure 3.4.

Forty-four studies collect multiple data, and three studies relied on a single sensor (GPS or Bluetooth) [54] [143] [155]. Moreover, there are distinct purposes for each data source. Locations are the significant context in social interactions. They can be gathered from GPS, Wi-Fi and cellular IDs. Messages and calls are two essential way of communicating through smartphones. Proximity is the primary condition that people can have face-to-face interaction, so Bluetooth is the most common method for sensing if two or more participants are in this situation due to its low power consumption and short distance signal. Microphone is capable of detecting surrounding sounds. By analysing raw audio captured from it, the condition if participants are engaged in conversations or not, can be inferred. Application usage, especially social media is an arising source of social interaction happened on smartphones which can not be ignored. One study even went further, it records Facebook connections and interactions [152]. Nevertheless, the aim of collecting other sensors data is not directly related to social interactions, but for studies' additional analysis. The accelerometer is for gaining participants' physical activities or step counts. The calendar is for participants' important daily life events. Smartphone usage is represented by quantifying screen on/off, phone lock/unlock, battery status and boot event. The ambient light sensor can provide surrounding illuminates readings. Music played on smartphones was especially monitored for personality researches.

Primary considerations for setting parameters of sensors are platform limitation and power consumption. In selected studies, ten of them described the parameters of some of the social sensors they used, including GPS, Bluetooth, Wi-Fi and microphone. Fifteen studies set fixed sample rates for recording data. GPS: 10 [12], 15 [236] [193], 20 [237] or 60 minutes [202] [199]; Bluetooth: 3 [35] [36], 5 [155] [57] [64] [253], 6 [137], or 10 minutes [237]; Wi-Fi: 6 [137], 15 [236], or 60 minutes [199]; and microphone: 3 seconds [237] or 3 minutes [28] [89]. [193] logged Bluetooth and Wi-Fi whenever the respective events occurred. Other three studies had dynamic sample strategies to balance battery impact and data quality [54] [53] [83]. For example, Kiukkonen et al's study [112] changed sampling rates depend on known Wi-Fi connections, motion and location status. If accelerometers and GPS showed participants are moving outdoors, it captured Wi-Fi data every 60 seconds and Bluetooth every 180 seconds. If smartphones were connected known Wi-Fi, it reduced Wi-Fi to every 120 seconds but increased Bluetooth to every 60 seconds.

3.4.4 Operating System

Thirty-five (77%) studies' data collection platforms are based on the Android system. Eight studies built on Nokia's Symbian OS. Six studies had their applications running on iOS. Two studies used Windows Mobile, and one study did not indicate which platform it was implemented on. In particular, five studies implemented their applications both on Android and iOS [45] [89] [90] [237] [241]. Symbian and Windows Mobile are popular in the last decade, and all selected studies based on them are before 2010. But these two mobile systems have stopped service nowadays. Compared with iOS, Android has flexible permission regulation, efficient background process and accessible sensor communications, which makes it easier to deploy data capturing platforms. So it is the most popular in selected studies.

3.4.5 Validation measures

Since there are different study targets of passive smartphone social sensing, various methods were applied for validating collected sensors data. Thirty-seven (79 %) studies adopted extra methods for affirmation, including professional clinical or psychological scales [182] [195] [28] [110] [74] [75] [89] [209] [193] [90] [66] [239] [240] [119] [35] [48] [152] [241] [208] [236], ecological momentary assessment (EMA) [239] [28] [45] [238] [239] [240] [89], on-line/smartphone-based surveys [57] [199] [125] [3] [66], self-designed questionnaires [155] [137], experience sampling method (ESM) [195], and in-person sessions [11]. As can be drawn from above, four studies implemented multiple validation approaches [195] [66] [239] [89].

For other studies, parts of the data themselves were regarded as validation. For example, One study considered weekly meetings of participants as the ground truth [53]. Based on these meetings, they could estimate the success rate of the smartphone of each person detecting other group members. In addition, Yu et al [253] separated data into learning sets and testing sets, it built a social network from learning sets and validate their model performance on testing sets. Bauer and Lukowicz [12] chose a stress period for students: before and after exam to observe their social behaviour changes. Similarly, Harari et al [88] monitored a whole academic term to characterise students' sociability.

In addition, these measures were taken at different time intervals. From studies describing the time point of administration of these measures, typically, professional clinical or psychological scales were applied once (at the beginning) by Faurholt et al [66] or twice (at the beginning and the end) by Harai et al [90] and Wang et al [239]. Particularly, Buck et al [29] had clinical assessments at three-month intervals, and two studies [240] [35] had their scales monthly. Pulekar and Agu[182] also applied psychological scales at the beginning and every 4 hours after. For EMAs, ESMs, and on-line/smartphone-based surveys, they were required multiple times a day [125] [239] [199], daily [45] [74] [137], every two days [238] [195], three times a week [28] [240] and monthly [3].

3.4.6 Data processing

Although all data were collected by applications running on smartphones, not all studies implemented their own tools for sensing. Twelve studies deployed platforms developed by others [28] [29] [66] [240] [48] [152] [241] [237] [153] [208] [45] [90]. Twenty-two applications (46.8%) specified that they used remote servers to store data transmitted from smartphones. Three studies (6%) stored their collected records on smartphones. Others did not describe how they aggregate their data. In addition, six studies created thresholds to filter the data. Two studies [90] [88] only included days with more than 14 hours data. Four studies [28] [238] [240] [241] set 19 hours as the minimum number of data needed per day. In Bati and Singh's study [11], they removed participants' data if their ground truth surveys were not complete or smartphones did not collect sufficient location data.

Ten studies (27 %) emphasised that their data collection procedures are ethically considered. Typical technical strategies are 1) limit the permissions of the application so sensitive information could not be recorded. For example, two studies mentioned that the content of messages can not be acquired in their sensing platforms [202] [11]. 2) anonymised identifiable entities such as IMEI numbers, Bluetooth/Wi-Fi MAC addresses and call/messages numbers [137] [253] [125]. These processes are usually done by randomisation and one-way hash. So the data can keep uniqueness but lose traceability. 3) encrypted secure connection when transferring data from smartphones to servers [239]. So the data cannot be intercepted or hijacked by unauthorised parties. Particularly, one study [112] strengthened that the rights of participants should be acknowledged. Participants should have the power to fully control their data. They designed a website for participants to view all their records and allow them to delete some or all.

Twenty-nine (62%) studies just utilised smartphones as raw data collection tools and did not implement any complex algorithms on smartphones. Their data analyses were conducted elsewhere afterwards. Otherwise, if a study collected microphone audio, it always processed the raw audio on the smartphone and stored the result. Twelve studies implemented algorithms to detect if the current sound is conversations or voices. Two studies recorded noise level of the surrounding environment. Furthermore, two studies [119] [114] derived types of activities such as walking, running from their application locally. Similarly, four studies [90] [238] [240] [241] utilised system built-in activity recognition interfaces including Google Activity Recognition [52] on Android and Apple Core Motion [51] on iOS to get activity types inference. Besides, five studies [238] [45] [239] [240] [119] combined the data from ambient light, microphone accelerometer and smartphone usage to determine if the participant is asleep. Precisely, if the smartphone is in a dark, silent environment, stay stationary and not being used, they inferred the user is sleeping. Lane et al [119] also considered recharging events for sleep since people often recharge their phones overnight. Furthermore, Wahle et al [236] established several thresholds to provide positive interventions from their smartphones. If participants stayed home too long, did not make any phone calls and walked less, recommend interactive activities will pop up to promote their mental state. Also, Lane et al [119] displayed animations to provide passive feedback to users based on their physical and social activities collected by smartphones.

3.4.7 Feature construction

Multiple methods range from straightforward numbers to complicated calculations have been employed for investigating collected data. All these tasks were conducted after the data collection period except studies giving feedback or intervention to participants during the experiment. Usually, if the study had validation measures, collected data will be interpreted and analysed with these ground truth to examine its hypothesis. For call logs, SMS logs, Bluetooth, Wi-Fi, conversations/voices, and smartphone usage, descriptive statistics were applied by reviewed studies. Total number, means, variation, standard deviations, and frequencies are calculated from plain numbers [182] [236] [202] [32] [195] [45] [75] [209] [90] [35] [36] [152] [241] [153] [199].

Besides, more features were formed by the characteristics of data. Calls were studied as long, short, incoming, outgoing and missed separately [182]. The difference and ratio of the number of these calls were also established as new features [202]. Similarly, messages were divided into sent and received, and length in characters is also recognised. Reviewed studies also extracted the number of contacts participants from call and SMS logs. For applications, they were evaluated based on their categories, which were selected manually [195] or according to the classification of Google Play store [75] [238]. Entropy which measures the diversity, unpredictability, or irregularity is also calculated for calls, SMSs, application usage, and Bluetooth [75] [209] [48] [152] [137] [237].

All studies build higher-level semantics from raw value instead of just observing these naive numbers for location data. Some features were formulated by the combination of the data and their temporal information. For example, in several studies, location points were clustered into interests of places by the length of time of each visit and frequency of visits [32] [202] [57] [143]. So significant positions such as home, workplace and socialisation venues were recognised by Tsapeli and Musolesi[228]. Features including time spent, distance travelled, number of unique places were calculated for further analysis. Moreover, all the features, including semantic locations were separated into different hours of days (for example, morning, afternoon, evening and night), weekdays, weekends to discover distinct patterns [57] [135] [110] [29] [110].

Further evolving processes were applied after features construction. Five studies calculated the correlation coefficient to select subsets of features to train machine learning classifiers [182] [11] [240] [237] [35]. Two studies [36] [202] selected certain features according to their predictive ability with the degree of redundancy [217]. In addition, three studies established their own scores on top of built features. Guo et al [83] designed the social tie matrix from calls, messages, Bluetooth and Wi-Fi records. Relaxation score was accumulated from touches of the screen, the number of messages and calls with different weight [114]. Lane et al [119] created well-being score from physical activity, sleep patterns and social interaction.

Table 3.1: Summary of reviewed studies.

Studies	Sample size & type	Study length	Data gathered	Purpose
Eagle and Pentland, 2006 [57]	100 students and staff in university	approx. 300 days	Call logs, Bluetooth, SMS logs, Application usage, Phone lock-/unlock, Cellular IDs	Recognize social patterns in daily user activity, infer relationships, identify socially significant locations, and model organizational rhythms.

Farrahi and Gatica-Perez, 2010 [64]	97 students and staff in university	approx. 300 days	Bluetooth, Cellular IDs	Proposed a model, which integrated the variations of location over multiple time-scales, and inferred interaction types from proximity.
Do and Gatica-Perez,2011 [54]	40 family members and colleagues	365 days	Bluetooth	Proposed a new probabilistic relational model to analyse long-term dynamic social networks created by physical proximity of people.
Oloritun et al, 2012 [163]	42 students in American university dormitory	approx. 300 days	Bluetooth	Understood the effect of social processes on the creation of social encounters at different lengths of interaction.
Do and Gatica-Perez, 2013 [53]	40 family members and colleagues	365 days	Bluetooth, GPS	Presented a probabilistic approach to mine human interaction types in real life.
Yu et al, 2013 [253]	30 student and staff working in the same building in university	approx. 300 days	Call logs, Bluetooth, SMS logs, Cellular IDs	Recognised human friendship from a supervised learning perspective, demonstrated the social relation evolution process by using the social balance theory.
Meurisch et al, 2015 [143]	163 students	28 days	GPS	Proposed a spatiotemporal approach to derive situational information about social interactions only based on location and time.

Pulekar and Agu, 2016 [182]	9 international students	14 days	Call logs, Bluetooth, SMS logs, Wi-Fi, Application usage	Proposed the Socialoscope, a smartphone app that passively senses user loneliness from their communication and interaction patterns, while factoring in different personality types.
Wahle et al, 2016 [236]	36 adults	14 days	Accelerometer, Application usage, SMS logs, Call logs, Calendar events, GPS, Wi-Fi	Explored the detection of daily-life behaviour based on sensor information to identify subjects with depression, and the potential of context sensitive intervention delivery.
Guo et al, 2016 [83]	38 family members and colleagues	approx. 730 days	GPS, Bluetooth, Call logs, SMS logs, Wi-Fi	Discovered three different types of geo-social behaviours, including online interaction, offline interaction, and mobility patterns.
Kostopoulos et al, 2017 [114]	5 young adult members from research group	30 days	Call logs, SMS logs, Wi-Fi, Application usage, Accelerometer, Screen on/off, Phone lock/unlock, Ambient light, Battery status	Proposed the StayActive, which using mobile sensor technology for detecting stress and recommend various relaxation activities “just in time”.

Bati and Singh, 2018 [11]	50, mostly age 18 to 21, education level 'some college'	70 days	Call logs, SMS logs, GPS	Proposed a new approach to model trust propensity based on long-term phone use metadata that aims to complement typical survey approaches with a lower-cost, faster, and scalable alternative.
Singh and Agarwal, 2016 [202]	54, 35 male, 19 female, most 18-21 years,education level 'some college'	70 days	GPS, Call logs, SMS logs	Described a novel approach to model an individual's cooperation level based on his/her phenotype i.e. a composite of an individual's traits as observable via a mobile phone.
Centellegher et al, 2016 [32]	142, age ranges from 28 to 50, 90 women, 52 men, 138 Italian, 4 from other countries	730 days	GPS, Call logs, SMS logs	Created a multi-layered view of the participants' lives, tracking social interactions, mobility routines, spending patterns, and personality characteristics.
Tsapeli and Musolesi, 2015 [228]	48 college students	70 days	GPS, Accelerometer, Calendar events	Discussed the design, implementation and evaluation of a generic quasi-experimental framework for conducting causation studies on human behaviour from smartphone data.

Madan et al, 2010 [137]	70 residents of an undergraduate dormitory	270 days	GPS, Call logs, Bluetooth, SMS logs, Wi-Fi, Accelerometer, Calendar events	Utilized co-location and communication sensors in smartphones to model the diffusion of health-related behaviours.
Moturu et al, 2011 [155]	54 from a community, which all members of the community are university related couples	210 days	Bluetooth	Explored the associations between sleep, mood and sociability by studying a population of healthy young adults going about their everyday life.
Schuwerk et al, 2019 [195]	234 adults	30 days	GPS, Call logs, Bluetooth, SMS logs, Wi-Fi, Application usage, Battery status, Music played, Boot events	Assessed autistic traits, social cognitive processing in everyday life and actual social behaviour.
Buck et al, 2019 [28]	45 patients	262.8 days	GPS, Call logs, SMS logs, Wi-Fi, Accelerometer, Microphone	(1) quantify between- and within-person variability in persecutory ideation (PI), (2) evaluate pre-existing models of indicators of PI, and (3) identify passively sensed indicators of PI.

Buck et al, 2019 [29]	45 patients	262.8 days	Call logs, SMS logs, Microphone	Evaluate whether smartphone-collected measures of social behavior can serve as early behavioral indicators of relapse among individuals with schizophrenia.
DaSilva et al, 2019 [45]	72 college students	72 days	GPS, Accelerometer, Screen on/off, Microphone, Phone lock/unlock, Ambient light	(1) Further the understanding of stress dynamics on college campuses by leveraging a dataset rich in passive sensing features to accurately, and naturally, capture possible stressors in the lives of students.
Khwaja et al, 2019 [110]	166 adults from 5 countries	21 days	GPS, Call logs, SMS logs, Accelerometer, Screen on/off, Microphone, Ambient light, Battery status	How do machine learning based personality assessment models perform across different countries? And what differences in the personality assessment models arise across different countries?
Fukazawa et al, 2019 [74]	20 adults	Not available	Application usage, Accelerometer, Ambient light	It proposed a method to predict the anxiety state of healthy people that combines these three features from smartphone log data.

Gao et al, 2020 [75]	183 college students and people working in the university	Not available	Call logs, SMS logs, Application usage	Proposed a multi-view multi-task learning approach with a deep neural network model to fuse the extracted features and learn the Big Five personality traits jointly.
Harari et al, 2019 [89]	152 adults	70 days	Call logs, Bluetooth, Application usage, Microphone	Examined the extent of: between- person variability in the daily assessments, mean level consistency across the daily assessments, and relationships among the daily behavioral tendencies.
Stach et al, 2020 [209]	624 adults	Not available	GPS, Bluetooth, SMS logs, Wi-Fi, Screen on/off, Phone lock/unlock, Battery status, Music played, Boot events	Examined the extent to which individuals' Big Five personality dimensions can be predicted on the basis of different classes of behavioral information collected via smartphones.

Schoedel et al, 2020 [193]	597 adults in the university context	30 days	GPS, Call logs, Bluetooth, Wi-Fi, Application usage, Battery status, Calendar events, Music played	Investigated how behavioural records from smartphones can be used to investigate individual differences in day–night patterns, how they relate to personality traits, and how they are influenced by intraindividual and interindividual factors.
Harari et al, 2020 [90]	633 college students	14 days	Application usage, Accelerometer, Microphone, Phone lock/unlock	Presented a conceptual framework and empirical illustration for personality sensing research, which leverages sensing technologies for personality theory development and assessment.
Wang et al, 2016 [238]	36 qualified patients	60 days	GPS, Call logs, SMS logs, Application usage, Accelerometer, Microphone, Phone lock/unlock, Ambient light	CrossCheck platform is the first step towards the passive monitoring of mental health indicators in patients with schizophrenia and paves the way towards relapse prediction and early intervention.
Faurholt et al, 2019 [66]	29 qualified patients and 37 healthy individuals	84 days	Call logs, Screen on/off	Investigated objective smartphone data reflecting behavioural activities to classify patients with bipolar disorder compared with healthy individuals.

Wang et al, 2014 [239]	48 college students	70 days	GPS, Bluetooth, Application usage, Accelerometer, Microphone, Ambient light	Showed a number of significant correlations between the automatic objective sensor data from smartphones and mental health and educational outcomes of the student body.
Wang et al, 2017 [240]	36 qualified patients	365 days	GPS, Call logs, SMS logs, Accelerometer, Screen on/off, Microphone, Ambient light	It was the first system capable of tracking schizophrenia patients' symptom scores using passive sensing and self-report EMA from phones, identify a number of passive sensing predictors of the clinical scores.
Lane et al, 2014 [119]	27 adults	19 days	Accelerometer, Microphone, Battery status	BeWell platform coarsely tracks the physical, social and sleep dimensions of well-being. Its feedback would allow users to easily understand the consequences of their actions.
Chittaranjan et al, 2011[35]	83 adults	240 days	Call logs, Bluetooth, SMS logs, Application usage	Showed that aggregated features obtained from smartphone usage data can be indicators of the Big-Five personality traits.

Chittaranjan et al, 2013[36]	117 adults	510 days	Call logs, Bluetooth, SMS logs, Application usage	Showed that significant relationships exist between personality traits and automatically aggregated smartphone usage cues.
de Montjoye et al, 2013 [48]	69 adults	Not available	Call logs, SMS logs	Showed that users' personalities can be reliably inferred from basic information accessible from all mobile phones and to all service providers.
Mønsted et al, 2018 [152]	636 college students	730 days	GPS, Call logs, Bluetooth, SMS logs	Predicted personality trait tertiles from a set of behavioral variables extracted from the data, and find that only extraversion can be predicted significantly better than by a null model.
Wang et al, 2018 [241]	646 college students	14 days	GPS, Accelerometer, Microphone, Phone lock/unlock	Used passive sensing data from mobile phones to examine the extent to which within-person variability in behavioral patterns can predict self-reported personality traits.
Bauer and Lukowicz, 2012 [12]	7 students	28 days	GPS, Call logs, Bluetooth, SMS logs, Wi-Fi	Described initial results from an ongoing project to use mobile phone sensors to detect stress related situations.

Madan et al, 2010 [137]	Not available	74 days	GPS, Call logs, Bluetooth, SMS logs, Wi-Fi, Accelerometer, Calendar events	Used mobile phone based co-location and communication sensing to measure characteristic behavior changes in symptomatic individuals.
Wang and Marsella, 2017 [237]	24 adults	70 days	GPS, Bluetooth, Wi-Fi, Accelerometer, Microphone	Explored behavior features extracted from smartphone sensing data, and used selected features to predict the traits of the Five Factor Model.
Harari et al, 2017 [88]	48 adults	70 days	Accelerometer, Microphone	Used a smartphone-sensing application to describe the patterns of stability and change that characterize a cohort of students' activity and sociability.
Montag et al, 2014 [153]	49 college students	90 days	Call logs, SMS logs	Linked self-report-data on personality to behavior recorded on the mobile phone.
Servia-Rodríguez et al, 2017 [199]	11000 users	90 days	GPS, Call logs, SMS logs, Wi-Fi, Microphone, Cellular IDs	Showed that these inferred routines are not independent from users' personality, well-being perception and other psychological variables. Explored predictability of users' mood by using passive sensing data.

Stachl et al, 2019 [208]	624 adults	30 days	GPS, Call logs, Bluetooth, SMS logs, Wi-Fi, Application usage, Screen on/off, Phone lock/unlock, Battery status, Music played, Boot events	Using a machine learning approach, it showed how these variables can be used to predict self-assessments of the big five personality traits at the factor and facet level.
LiKamWa et al, 2011 [125]	25 adults	21 days	GPS, Call logs, SMS logs, Application usage, Calendar events	Showed that user mood can be inferred into four major types with an average accuracy of 91%.
Aharony et al, 2011 [3]	185 adults	510 days	Call logs, Bluetooth, SMS logs, Wi-Fi, Application usage, Accelerometer, Screen on/off, Battery status, Cellular IDs, Application usage	Introduced the Friends and Family study, a longitudinal living laboratory in a residential community.

3.4.8 Data analysis

After abstracting features from plain data, correlation analysis or machine learning algorithms were usually implemented in the next step to explore the relation between ground truth and collected smartphone data. Sixteen studies (34%) performed correlation analysis, such as Pearson's correlation [11] [32] [155] [238] [239] [240] [35] [36] [237] [153], Spearman correlation [236] [89] [90] [193], Jaccard similarity coefficient [32], Kendall correlation [228], and test-retest correlation [88]. Also, six studies calculated different kinds of coefficient, including intraclass correlation coefficient (ICC) [89] [90], within and between participants coefficient [28], regression coefficient [29], generalised estimating equations (GEE) coefficient [240] and correlation matrix [45]. In particular, one study [119] used Levenshtein similarity to compare its self-created scores and survey results, and one study [135] applied the Phase Slope Index (PSI) to measure temporal information flux between time-series signals.

Typical machine learning algorithms utilised by reviewed studies are Support Vector Machine (SVM) [253] [236] [110] [35] [48] [152] [237], Random Forest [143] [236] [182] [11] [202] [110] [74] [209] [208], regression analysis [202] [137] [208] [238] [66], AdaBoost [182] [11] [202], Naive Bayes [182] [202] [110], Neural Networks [75] [199], Bayes Net [182] [202], Probabilistic Model [54] [53], Hidden Markov Model [57], Gaussian Mixture Model [57], Latent Dirichlet Allocation [64], Exponential Random Graph model [163], Decision Tree [182], Gradient Boosted Regression Trees (GBRT) [241], KStar [11], LogitBoost [202], and XGBoost [74]. Five studies attempt different machine learning methods, compared their performances, and chose the best alternative[143] [182] [236] [11] [202]. When comparing, two studies also feed particular sets of features (demography only, smartphone only and both) to the algorithm [11] [202], and the results showed that smartphone based features all outperformed. Four studies [209] [110] [74] [241] also implemented cross validation to their machine learning models to eliminate over-fitting. Moreover, mean absolute error (MAE) and root mean squared error (RMSE) are usually adopted to evaluate the performance of the model [75] [240] [241] [88] [208].

3.4.9 Benefits

Almost all studies illustrated the reasons for applying passive smartphone social sensing and discussed the benefits of it. The most prevailing incentives are: 1) Ubiquitousness and unobtrusiveness: Almost every people have a smartphone nowadays, and it is natural to carry smartphones and use them for communication habitually [57]. Measures using smartphones do not ask participants to carry extra devices which may interfere with their normal behaviour [53]. 2) Capability and continuity: smartphones are equipped with various sensors [137], and they can monitor both contextual and behavioural information of participants without interruption over a long period. So researchers are able to observe changes and deviations from a comprehensive perspective [11] [202] [32] [193]. 3) Personalisation and individualisation: all collected smartphone data are from the exact participant, so it provides researchers chances to construct in-depth models for this individual. This is particularly important for health-related studies because dedicated treatment or interventions can be introduced [28] [88]. Specifically, one study strengthened the advantage of Bluetooth for proximity detection, which includes low battery cost, high compatibility in distinct environments, popularity among devices, and less-privacy sensitive compared with voice and location [53].

3.4.10 Problems and challenges

The counterbalanced issues and challenges are brought together by benefits of passive smartphone social sensing. These problems were involved in different stages of the study, including data collection and analysis. They were summarized in the following three categories.

Privacy

Protecting sensitive information of participants, especially identifiable part, is always the highest consideration for almost all studies. However, only a few of them describe how these procedures were handled in detail. Usually, user identity such as phone numbers, MAC address and IMEI were hashed irreversibly to be anonymized before analysis [253] [11] [137]. In two studies [11] [202], the application was specially designed for requiring fewer permissions than common ones. Centellegher et al [32] developed a digital space, which participants can control and disclose their own data. Some studies also suggested particular methods for participants' privacy, such as ignoring individuals but building coarse-grained systems [253], sharing only statistical summaries, and inserting random perturbations [148]. For studies giving phones to their participants, the privacy consideration even threatened the validity of the experiment. Buck et al [29] reported that some participants refused to use given phones as their primary ones because they were aware that their activities are tracked.

Accuracy

Although smartphones are with participants almost everywhere, we can not fully regard smartphones as users' themselves. Participants may break, loose, and neglect to use or charge their phones [240]. Consequently, Eagle and Pentland [57] implemented a forgotten phone classifier by observing if the smartphone was charging, staying in the same place for a long time, reaming idle through missed calls, messages. Most studies rely on Bluetooth as indicators for face-to-face interaction. Nevertheless, it is not the original intention of this technology, and it has its own technical defects. Some studies reported it could not detect all nearby device in a scan and quite noisy [57] [54] [53]. Similarly, it is almost impossible to examine other inferences such as conversations, locations and activities. Since there are a large amount of data, the collection of ground truth would be too disruptive for participants [119]. Besides, due to different sensor limitations on Android and iOS platforms, collect data from two platforms can not be merged correctly. So further data analysis had to be applied, which may destroy the coherence of the results [241] [89].

Methodology

The actual world is always much more complicated than our assumptions. Various problems could appear under passive smartphone social sensing. One study[253] reported that some participants did not have any calls or message during the experiment period, which could because the time of study is not long enough. As stated in the accuracy problem, face-to-face interaction is usually inference from Bluetooth signals are within transceivers' range, but it does not imply that participants are in any form of interaction necessarily [163]. Moreover, as mentioned in the participant sections, only one study reported how the number of participants was decided. Most participants in reviewed studies are related to universities. They are either students, staff and researchers or families, friends and people living around them. The homogeneity of participants, uncontrolled study design combined with small sample size threatens the generalization of experiment results, which is the most common concern among all studies [253] [202] [228] [137] [155] [236] [152] [237] [88]. In addition, the ground truth almost all studies relying on are mostly self-assessment-base scales. They have natural deficits such as subjectivity and recall bias, which means they are not the perfect gold standard [74].

3.5 Discussion

3.5.1 Findings from reviewed studies

Reviewed studies illustrated the existing utilisation of passive smartphone social sensing. Although most studies only employed this technique as a novel instrument to investigate human social behaviour, we can still notice its potential applications in health-related researches, such as mental health, depression, sleep, etc.. Moreover, comprehensive procedures of passive smartphone social sensing, including study design, data collection, storage and analysis are exemplified, which provides interested computer science, social, and psychology researchers beneficial references. Besides, the continuity and unobtrusiveness of the smartphone offer more precise and in-depth monitoring without imposing burdens on participants. Reviewed studies generally demonstrated the potential and capability of passive smartphone social sensing for reflecting people's social behaviour. For example, phone-based models had significantly better performance than traditional demographic models in Singh and Rishav's study [202]. Smartphone passive social sensing is a promising technology in the filed not only because of its non-intrusiveness and unobtrusiveness, but also smartphones are indeed the hub of personal communication. Phone calls, messages, usage of social media applications can not be ignored when measuring social interactions. In addition, although not utilised as their original purpose, variety of sensors embedded in smartphones enable the context and environment information retrieval. However, studies are still necessary to confirm the hypothesis that, passive smartphone social sensing is more accurate and efficient but less interruptive and troublesome than existing measurement.

Sensing strategy

All reviewed studies except those were analysing existing datasets implemented particular applications on smartphones for data collection. But only a few detailed reported within the study or elsewhere about, how decisions such as, which sensors to monitor, frequency of sensors were made during the development. For the few who did, the principal consideration is battery consumption. For example, the data collection campaign [112] analysed in these three studies[54] [53] [83] selected optimising power consumption as the basis of application development. It turned the sampling rate according to the condition of smartphones such as mobile/stationary, connected to known Wi-Fi. Another example, Meurisch et al [143] implemented Karken [196] as data collection framework, which was constructed on a greedy approach. It gathered as much data as acceptable, considering only privacy and energy consumption. However, none of the reviewed studies chose the parameters of sensors from the purpose of their study perspective. Indeed, higher granularity data provide more possibility of better model performance intuitively and in practice [37]. But simultaneously, the expense it brings, including the privacy, transmission and storage dilemma can not be neglect. For example, Eagle and Pentland [57] exhibited data corruption during the collection period. It was caused by continuously writing data from sensors and the finite number of read-write cycles of flash memory card. Nearly a month of data of certain participants was lost. Although no other reviewed studies reported issues of collecting, transferring and storing huge amount of data the passive smartphone social sensing produced is still challenging. It is worth attention for researchers to determine how much data they actually need according to the actual competence of the device and the purpose of the study.

Privacy

This trade-off is also applicable to the privacy of participants. Although no participants in reviewed studies complained about their unacceptable experience and violation of personal information, collecting such a high volume of sensitive data on smartphones will certainly involve privacy issues. All studies claimed that they have appropriate ethical approvals and consents from participants, but none of them reported the struggle during that process. So it is still questionable that, are data collected from existing sensors excessive or deficient in confirming hypotheses in terms of privacy? From all reviewed studies, the most privacy-sensitive part in passive smartphone social sensing only involves identities of call, messages, social media, Bluetooth, and length of verbal conversations [239]. But the content of these social contacts can provide more comprehensive knowledge of participants' social behaviour, especially for those investigating mood, sentiment, mental health and related disorders [6]. Such data will contribute to a more compelling conclusion potentially. But it may raise a severe dispute in the ethical committee.

Besides, participants may have various attitudes towards different types of sensors. They valued such data, but not the same across all sensors [10]. For example, Predrag et al studied 24 participants attitudes towards personal sensing [113]. The results show no attention was given to the accelerometer and barometer, but concerns were highly grown for sensitive ones such as microphone and GPS. Participants considered GPS data is 'creepy' and could threaten their physical security. Nearly all of them had negative attitude towards raw audio. They felt 'too watched and too listened to'. However, recording audio at the necessary frequency for activity inference was more acceptable. Participants in the study also have different concerns on the length of time the data were kept. In general, raw data is unwilling to be kept for both GPS and microphone. As long as the inference was accomplished, the raw data should not be retained. The living context of participants also influences their preferences. If a participant has to share sensitive information at work, the audio recording is definitely not welcomed [93]. The value of collected sensor data plays another role in deciding the acceptability of sensors. For example, a runner would like to know his workout performance, so raw GPS data would be likely to be kept for a longer time for analysing routes, pace and distance [113].

The knowledge of sensors' capability could influence the attitudes towards each sensor. Nguyen et al found privacy problems are highly concerned by participants. But these considerations are only on the abstract level, actual everyday tracking technologies such as RFID and web records are reported significantly less concerned [158]. It's probably because rather than specific sensors terms, participants are more familiar with descriptions used in their daily lives. Furthermore, the understanding of sensors could shape users privacy concerns. With higher level understandings of the implemented technology, more worries could be raised by participants [16]. Nevertheless, there is also a study that found users who already had sensor abled devices are more willing to adopt this monitoring technology [158]. Sharing preferences of collected sensor data also have a hierarchy for different types of contacts. Participants shared more sensor collected information with strangers than their own family and friends in Prasad et al's study [177]. If specific third parties provide enough benefits, participants were more willing to share. The study also suggested users' privacy concerns are not static, and sharing decisions could be changed over time [177]. Moreover, necessary privacy strategies could be applied using the preferred technology by target users. An appropriate interface established for users to manage privacy is a reasonable start. Christin et al tested six graphical privacy interface for 80 participants. But found there is no universal preference of the majority of participants [38]. Users favoured elements with different colour and size to visualize the privacy protection level and define their preferred privacy settings. So user ability tests could be conducted before the actual experiment to determine the suitable privacy interface.

In general, decisions of achieving the best results with minimum violation of participants privacy have to be made by researchers. Basic knowledge, possible concerns and potential benefits of each sensor could be informed to participants to help them made their own decisions. All factors including the demographics, attitudes, aims of studies should be considered to apply appropriate sensing strategies for the particular population of participants. Participants' preference for sharing particular kinds of sensors data with different types of people should be respected.

Sample size and integrity

All reviewed studies described sample size and demographics of participants to some extent, but they did not demonstrate any significant clinical value if they are health-related research. Although the small sample size is satisfied for feasibility or exploratory studies, the accuracy and precision of the statistical results are hampered substantially [55]. The majority of participants are still college students, researchers or people related, which also deteriorate the generalisation of the study. Usually, small sample size studies provide opportunities to enhance data integrity [121], but only one of the reviewed studies demonstrated the completeness its data which is 85.3% [57]. From strategies which other studies utilised to filter the data, for example, counting only the day with at least 15 hours' data as a day to analysis [88], it can be realised that 100% acquisition of data in passive smartphone social sensing is not always applicable. It could be caused by different reasons such as application corruption, sensing platforms, storage errors and phone turned-offs. Although no studies reported how data integrity influence the final results, the data loss is always a hidden problem and possibly affect the quality of analysis. Consequently, researchers should have convincible justifications of sample size choice and adopt measures to minimise errors that happened during the data collection.

3.5.2 Implications for future studies

Sensing strategy

From reviewed studies, various configurations of sensor application, including types, frequency, and combination, were demonstrated. Nevertheless, the energy consumption are major concerns why reviewed studies deployed these configurations. Certainly, specialized designed processors added on smartphones recently makes collecting high-frequency data more energy-efficient [43]. Limitation considerations such as platform restrictions, power consumption and participants' privacy are still necessary. But do these constraints are the only reason researchers made that choice? Do we demand all sensors in that high frequency to reach these limitations? Or are lower sampling rates enough to achieve good results? Indeed, the chosen sensor parameter of reviewed studies could answer their research questions. But it's from the results of these studies rather than theoretical background or empirical experiments. From the reviewed studies, the procedure of settings sampling rates seems instinctive. Most studies just illustrate the parameters of sensors plainly. For example, section 3.4 summarised that the GPS frequencies in the reviewed studies are 10, 15 or 60 minutes. These numbers are given directly by reviewed studies without many variations. Their primary consideration is battery usage. However, sensor data frequency may influence these results. Different resolutions could provide better performance or cause more interference. There still room for balancing power consumption and sensor frequency. By applying a more efficient sensor frequency, a better outcome with lower battery usage could also be achieved. These questions are fundamental in smartphone passive social sensing and need additional investigations.A nontrivial method for constructing the sensing strategy could be initiating it from the goal of the study, which may alleviate these issues from the beginning. So a more reasonable choice of sensor shall be made and better results could be achieved. Experiments on different sample rates and combination of sensors could be conducted to examine which choice is the most efficient in achieving the study's purpose. Although there exist studies trying to clarify these issues such as [131], which explored the necessary Bluetooth signal strength for inferring face-to-face proximity in various situations, there still plenty of unsolved problems in this field. For example, for recovering a certain percentage of the face-to-face interaction of participants, how frequent the Bluetooth scan should activate, how the frequency of Bluetooth scans affects the accuracy of face-to-face recovery? With these verification studies, researchers can make legitimate decisions on which sensors to capture and their sampling rate at the planning stage of the study. These experiments will also contribute to clinical value for health-related study and lay a solid foundation for sensor usage of passive smartphone social sensing.

Causation and personalisation

Although reviewed studies have examined the feasibility and validity of passive social sensing, the most commonly reported results are correlation coefficient and performance of applied machine learning algorithm without precious causation explanation. Observational research usually observes individuals directly in natural settings. Therefore, for cohort studies, alternative explanations for results due to confounding could exist [31]. Moreover, researchers may only focus on the designated variables and ignore other possible factors. Two main methodologies have been proposed to control such defects: structural equation modelling and quasi-experimental designs. The former one applies multivariate regression, and the second one used a matching design to exploit inherent characteristics of observed data [228]. There exists research using collected smartphone features to conduct a quasi-experimental study which shows the potential for causal studies. Comparing with demographics, smartphones introduce plenty of additional confounding variables, which need to be specially considered. Intuitively, features such as the number of phone calls and messages can reflect participants social interactions to some extent. Some studies also have utilized other knowledge to formulate higher-level features; for example, diversity of calls, messages and GPS based on Shannon Entropy were created [202]. These features are applicable for observation studies, which gives researchers implications and directions for their future work. But strong correlations or classifications with these features do not indicate that they are reliable measures, especially when the demographics and sample sizes are restricted. These issues may threaten the generalisability of the results and applicability of smartphone passive sensing. For example, there is a discussion in a reviewed study of personality that findings do not match well with previous results [208]. It attributed the difference to the type of data used. Further theoretical investigations or cross-population experiments could be considered based on current passive smartphone social sensing results. Smartphone features could be treated as items in questionnaires to be validated and reasonably interpreted. Variables constructed from smartphone data could be timely standardised. For example, the number of calls could be calculated and reported by weeks or month. So cross-compare among different studies will become achievable. Then passive smartphone sensing could evolve from a promising field to a practical instrument. So the data generated by smartphone could be more instructive rather than just possible correlations.

It could expand the smartphone passive social sensing into a wider filed, clinical studies. Using these kinds of sensing technology including smartphones, wearable device and in-home monitoring often termed digital phenotyping in that area. It refers to 'moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices' [164]. Digital phenotyping has been applied in various disease researches, such as mood disorder [26] and schizophrenia [227]. They all suggested digital phenotyping is actionable and potentially useful in future clinical outcomes. However, to our knowledge, none of the digital phenotyping technologies has been approved for clinical usage, and few, if any, have been adopted to replace traditional health monitoring [42]. Studies utilising digital phenotyping are often small-scale, coarse and unstandardised [65]. So they are insufficient for effective analysis and not suitable for robust identification of clinical signals [98].

To reach clinical validity, different factors contributing to smartphone usage have to be considered for experiment design. As mentioned before, sensing platforms favour Android as it has fewer constraints. But it excludes a large number of other platforms users such as iOS. Different age groups may have distinct smartphone interaction patterns. Younger generations who are used to having the smartphone may spend much more time on it than the elderly aged. The assumption that smartphones are always carried on" may not be applicable in older populations [98]. Machine learning algorithms could also exaggerate this bias. By training data from limited populations, the model and results could overfit those groups of people, which can not be derived for wider communities. However, it doesn't mean digital phenotyping can not be used clinically at all. With the continuity of smartphone sensing, longitudinal observations rather than snapshots could be provided to clinicians. So digital phenotyping could be used to explore the mechanisms and behaviours underlying psychiatric disorders rather than outcomes alone [150].

In addition, although all conclusions drawn from reviewed studies are in population levels, some researches have claimed that applying personal models are more efficient than population-based ones [62]. There is also evidence showing that each individual may have unique social patterns revealed by particular characteristics [43]. It is especially effective in mental health-related studies because dissimilar behavioural indicators of mental health difficulties were found in different people [17]. It mirrors N-of 1 approaches which argued that should offer better efficacy than one-size-fits-all [181]. Therefore, another promising field of passive smartphone social sensing is the personalisation of the model. Each participant can be treated as a singular case, so different features will be analysed to identify which of those influence the most. Various cases could also be cross-compared with validation measures to explore their exceptional patterns. The personalisation is not that strict to a single person. Similar characteristic such as age, gender and personality could be grouped together to generate 'similar user' models [183]. These models will provide opportunities to discover how these characteristic affect social behaviour. It can enlarge the volume of data for specific algorithms if the data from a single participant is not enough [43].

The next phase

Since the inclusion criteria require all reviewed studies are empirical, so not all of them have formal sensing platforms or their platforms were reported elsewhere. So a fully systematic summary on these platforms can not be achieved. But every sensing platform should have documents, instructions and give opportunities for other researchers to use. It could save plenty of time for repetitive development so researchers could deploy their sensor strategies easily on these platforms. As engineering development, smartphone sensing platforms rapidly advanced over time [245], and an academic review could be done on these platforms. So researchers could know the specifications and choose the appropriate one for their studies. It also applies to feature extraction and machine learning approaches, standard and unifying approaches would boost the analysis of collected data and communications across the community.

Dedicated chips have been implemented to manage the power consumption of embedded sensors, which empowers higher sampling rate but less battery drain. The computing power and storage capacity of smartphones has also increased exponentially in a decade. The emergence of bionic processor and machine learning models such as Apple's Core ML and Google's TensorFlow enabled developers to deploy complex algorithms on smartphones. So the data could be processed locally without transferring sensitive information to remote servers [43]. All these advance of technologies will bring great relief for all privacy concerns because it is not necessary for other people to access real private data. The whole circle of collection, analysis and removal will be accomplished on smartphones and controlled by participants themselves.

However, problems arise simultaneously can not be ignored. The stricter privacy considerations of smartphone operating systems bring challenges of passive social sensing platform. For example, Google emphasized that the accessibility service in Android, which most sensing platform utilized to collect data, should 'only be used to assist users with disabilities'. It caused many sensing applications like AWARE [67] have to leave the Google Play Store and distribute on their own, which increase the difficulties of deployment. In addition, the popularity of social media applications changed the role of smartphone communication. People have switched the channels from traditional phone calls and messages to video chatting and social media messaging [169]. But unlike original ones, there are numerous social media applications which participants may use. Only a few of them provides public application program interface (API) for developers to query social interaction data under the acknowledgement of users. So researchers can not obtain the full picture of participants' social media interaction. The usage of social media also involves new methods of communication such as like, comment, repost, etc.. How doses these new methods affect overall social behaviour still need further investigation.

3.6 Limitations

Although the search string considers different variations of the smartphone, the full coverage of terminology is not guaranteed. The two emphasized terms social and sensing may not include particular results. Not all possible fields passive smartphone social sensing have been covered. The number of included studies is relatively small, even from an extensive search because of strict inclusion criteria. Studies describing and discussing the procedure of data collection, implementation of platforms, related algorithms and privacy theories were excluded because they did not involve any practical experiments. Mobility, proximity and context-aware studies were also discarded due to their main aims, which is not social sensing. But these papers are still valuable and beneficial from other perspectives for the field. Smartphones with other wearable sensors for passive social sensing is an efficient complementary for smartphone sensing alone. With specialised sensors, it can capture dedicated types of data such as heart rates and sleep patterns. Since the scope of the review, we didn't include that in the review.

3.7 Conclusion

Social behaviour is a significant component of human behaviour, and it has been confirmed to be correlated to many other psychological, and health-related factors such as mood [155], stress [114], and depression [225]. Passive smartphone social sensing provides a novel opportunity other than traditional psychology questionnaires to probe the insight of human social behaviour. Although this technology has been used in empirical studies, its key components, such as performance, strength, and limitations, have not been methodologically reviewed. The systematic review discussed a series of questions about smartphone passive social sensing. Fundamental resolutions include how to gather social interactions on smartphones and how to infer social interactions outside smartphones were inferred were presented. It is the first time that the whole procedure and technology of passive smartphone sensing was systematically summarised. In general, calls, messages, and social media usage are three major sources of social interactions on smartphones. Bluetooth microphones could be used to infer social activities outside the smartphone, such as face-to-face conversations. All these data sources need to be constantly observed to give a full picture of participants' social behaviours. This systematic review will be a practical reference for researchers applying this technology to related social behaviour studies. It will be beneficial to all relevant researchers because it has the competence to be employed in various social interaction-related areas such as colleague cooperation, teaching performance, and political opinion propagation [136]. All gathered data are individualized, precise and objective, which could inspire an in-depth understanding of the phenotypic social behaviour of each individual. They will also empower precision feedback or intervention if necessary. However, to achieve these ambitions, issues such as theoretical basis, privacy policy, and experiment significance should be further explored. Moreover, it is essential to keep track of the evolvement of people's social interaction habits, such as the appearance of new communication channels. To address the gaps between present state-of-the-art and the vision, interdisciplinary collaborations between technology experts, computer scientists and psychologists are required. Through these actions, passive smartphone social sensing could be the standard social behaviour measurement in the future.

Chapter 4

Understanding Individuals' Compliance with COVID-19 Policies

The systematic review exhibits the feasibility and originality of passive smartphone social sensing. Following the discovered paradigm, we adopted this technology to actual PD participants. After the ethics committee approved the whole experiment plan, the participant recruitment started in September 2019 ended in March 2020. A sensing application was installed on participants' smartphones for collecting their social behaviour data 27/4 for one year. Unfortunately, COVID-19 had become a worldwide pandemic since then, and the UK government introduced a series of policies to reduce the transmission of the disease. People had to stay at home, keep social distancing, and maintain a social bubble. All these restrictions dramatically impacted the regular social life of participants. However, it gives an opportunity to examine how our monitoring reflects social behaviour changes caused by COVID-19 and participants' obedience to these policies. Based on the assumption that one Bluetooth signal could represent a person around, the number of one time and unique signals could represent the number and variety of people around. GPS data can also detect the time participants spent at home. So, their reactions towards these restrictions are revealed. The capability of our smartphone monitoring method is also examined in this chapter

The content of this chapter is adapted from *Ibrahim, Ahmed, Heng Zhang, Sarah Clinch, Ellen Poliakoff, Bijan Parsia, and Simon Harper. 'Digital Phenotypes for Understanding Individuals' Compliance With COVID-19 Policies and Personalized Nudges: Longitudinal Observational Study.'* JMIR Formative Research, May 2021. Volume: 5. ISSN: 2561-326X. DOI: 10.2196/23461.

Author's contributions

Heng Zhang designed and conducted the GPS, and Bluetooth data analysis of Parksinons' participants, summarised the results, discussed the findings and wrote a significant part of the manuscript. He is the primary author of section 4.2 and 4.3, joint author of section 4.1, and provide secondary input in section 4.4 and 4.5. Ahmed Ibrahim designed and conducted his participants' GPS and application usage analysis, summarised the results and discussed the implications for personal nudges. He also wrote a significant part of the manuscript. Ellen Polikoff, Bijan Parsia, Sarah Clinch and Simon Harper provided the guidance of the paper and gave suggestions on the writing of the manuscript.

Abstract

Background: Governments promote behavioral policies such as social distancing and phased reopening to control the spread of COVID-19. Digital phenotyping helps promote the compliance with these policies through the personalized behavioral knowledge it produces.

Objectives: This study investigated the value of smartphone-derived digital phenotypes in (1) analyzing individuals' compliance with COVID-19 policies through behavioral responses and (2) suggesting ways to personalize communication through those policies.

Methods: We conducted longitudinal experiments that started before the outbreak of COVID-19 and continued during the pandemic. A total of 16 participants were recruited before the pandemic, and a smartphone sensing app was installed for each of them. We then assessed individual compliance with COVID-19 policies and their impact on habitual behaviors.

Results: Our results show a significant change in people's mobility (P¡.001) as a result of COVID-19 regulations, from an average of 10 visited places every week to approximately 2 places a week. We also discussed our results within the context of nudges used by the National Health Service in the United Kingdom to promote COVID-19 regulations.

Conclusions: Our findings show that digital phenotyping has substantial value in understanding people's behavior during a pandemic. Behavioral features extracted from digital phenotypes can facilitate the personalization of and compliance with behavioral policies. A rule-based messaging system can be implemented to deliver nudges on the basis of digital phenotyping.

4.1 Introduction

4.1.1 Background

COVID-19 is a highly contagious disease with confirmed cases in more than 188 countries as between December 2019 and June 2020, resulting in a global pandemic [244]. To control the spread of COVID-19, governments have enforced behavioral policies, such as stay-at-home and social distancing measures, which limit the usual patterns of human interaction [5, 27]. The potential risk of problems with social isolation [231] complicates the implementation of these policies, which places an additional responsibility on governments to maintain mental health throughout the pandemic.

Currently, governments rely on communication campaigns to persuade people to adhere to COVID-19 behavioral policies and reduce disease spread. Health agencies, such as the National Health service (NHS) in the United Kingdom, design communication in a way that encourages the application of the promoted behaviors while avoiding problems related to social isolation. This approach to communications design employs behavioral insights derived from scientific studies to deliver behavioral guidance [232]. The communications resulting from this process are called “nudges” [122].

Despite the critical role of these campaigns in elevating community awareness, they are not designed to reflect differently when people exhibit different behavioral responses to the promoted procedures. Digital devices including smartphones can be used to recognize behavioral differences. Accordingly, communications can be personalized and contextualized on the basis of the individual’s behavior. Smartphones facilitate the capturing of behavioral features through the continuous and unobtrusive collection of sensor and interaction data; this process is known as “digital phenotyping.”

In this study, we show how an individual’s behavioral reactions to COVID-19 policies can be observed through digital phenotyping. Subsequently, we suggest a personalized way of delivering nudges designed around the individual’s reactions to the enforced regulations. We report 2 longitudinal studies that started before the outbreak of the pandemic to collect digital phenotypes. Our studies allow us to observe the impact on the overall behavior before and during the outbreak. Additionally, we observed the impact of COVID-19 on habitual behaviors and the uptake of new apps.

Our primary research contribution is the introduction of an approach that employs behavioral differences derived from digital phenotyping in the design of personalized nudges. Although we did not conduct an experiment to measure the real-time effects of personalized nudges, the proposed nudges conform to the general guidelines in behavioral science and are expected to improve individual compliance to them. Moreover, the development of mental health issues as a result of lockdown policies can be observed through digital phenotyping and better addressed through personalized nudges.

4.1.2 Related work

With the popularity and evolution of personal electronic devices, people are producing an increasing number of digital footprints such as those generated through web-based communication and mobile device usage. These footprints can be linked and analyzed with clinical data to create an individualized, nuanced view of human disease, which is called a "digital phenotype" [100]. In 2015, a digital phenotype was defined by Jukka-Pekka Onnela as the "moment-by-moment quantification of the individual-level human phenotype in-situ using data from smartphones and other personal digital devices" [226]. Digital phenotyping has become one of the most innovative approaches to enhance health and wellness via human-computer interactions through digital technology.

Nowadays, smartphones have become the one of the ideal tools for digital phenotyping. Smartphones are the hub of personal communication, and almost everyone has a smartphone. Although smartphones are not specially designed for behavioral research, they can collect a large amount of related data directly and instantly with ecological validity. Social interaction on smartphones, including calls, messages, emails, and social media usage, can be captured without difficulty. Thus, social sensing could be less intrusive on smartphones than on any other device. Embedded multiple and power sensors also empower smartphones as an efficient tool to record the surrounding social context. For example, raw data from sensors such as microphones, the global positioning system (GPS), and accelerometers can be gathered and interpreted as conversation engagement, mobility patterns, and the number of encounters to infer social interaction occurring outside of smartphones. Thus, smartphones could be one of the most applicable ways of passive societal digital phenotyping.

Digital phenotyping on smartphones has been utilized in various fields, especially psychological and health-related studies. Abdullah et al [1] collected phone usage patterns to detect and predict discrepancies in sleep rhythms. Furthermore, LiKamWa et al [126] analyzed call, message, or email contacts and location clusters from smartphones to infer users' daily mood. Farhan et al [63] combined the locations and activities from participants' smartphones to predict depression. Boukhechba et al [23] explored the association of social anxiety with GPS and communication patterns. To confirm the findings and observations of passively collected smartphone data, all these studies asked for participants' input through various means including interviews, focus groups, and questionnaires. All these studies claimed to have relatively high accuracy. Albeit with different aims, our study similarly implemented these smartphone monitoring technologies. We collected data before and during the COVID-19 lockdown, which provided us an opportunity to observe individual behavioral changes. We also conducted interviews with our study participants to verify our findings.

4.2 Methods

4.2.1 Methods Overview

We used behavioral indicators for the COVID-19 policies as proxies that would help us observe the adoption of the desired change by people. Our approach relies on transforming raw smartphone data collected longitudinally (ie, digital phenotypes) into behavioral features. Distance travelled and time spent at home by a person are examples of features derived from raw location data (ie, timestamped longitude and latitude attributes). The detection of behavioral indicators is achieved at the level of behavioral features rather than the raw data. This is because behavioral indicators are manifested at a higher level of human understanding expressible by those features. In the following section, we detail the behavioral features and their roles in recognizing the behavioral indicators of the proposed policies.

For this disease, transmitted through close contact, reducing the possibility of an uninfected person having physical contact with an infected person may be the only effective way to suppress the transmission of the disease. Since the onset of the COVID-19 pandemic, governments worldwide enforced a series of behavioral policies based on this concept to control the spread of this highly infectious disease. For example, the government of the United Kingdom instructed individuals to stay home as much as possible, to limit contact with those from other households, and to maintain distance from others when stepping out of home (2 meters apart where possible) [77]. Other measures include school closures, working from home, cancellation of mass gatherings, and travel restrictions. These policies are referred to as "social distancing" or "physical distancing" policies.

4.2.2 Stay-at-Home Measures

Deriving behavioral indicators of social distancing from smartphone data was our primary consideration. There are some existing studies on the mobility responses to COVID-19; for instance, a previous study [249] analyzed public geolocated Twitter data to measure the travel behaviors of users. Allcott et al [4] combined surveys and GPS foot traffic patterns to observe partisan differences in social distancing. They reported a substantial reduction in the mobility of people in the United States, albeit with partisan gaps in beliefs and behavior. Similarly, we can expect that our participants should spend almost all their time at home and to limit the time and number of places when stepping out, which is usually only for essential shopping owing to the implementation of social distancing measures. These behavioral changes can be acquired from raw GPS data. Since participants' smartphones record latitude and longitude attributes continuously, their distance from home can always be calculated. Thus, we can determine the time and frequency of their trips outside of home.

Furthermore, social distancing measures can bring about adverse effects, especially on mental health. Some of these reported effects include stress, anxiety and depression, and panic [70]. To maintain mental well-being and while at home, people may find alternative methods of communication to replace their regular face-to-face interactions. Phone calls, messages, video chatting, and social media are possible substitutions people may choose; accordingly, a potential increase in the use of these communication methods is expected. With the various data sources, we could draw a comprehensive and personalized picture of how people react to the impact of COVID-19 restrictions.

4.2.3 Social Distancing Measures

Social distancing implies that people should meet fewer people than they would during normal times. Bluetooth signals are an effective reference for face-to-face interaction recorded on smartphones. Nowadays, almost everyone carries a smartphone, and almost every smartphone is equipped with Bluetooth technology, which scans surrounding signals and reports its identity continuously in a short range. Thus, every newly captured Bluetooth entry could potentially represent a new person in close proximity [131]. This technology has been wildly used in the field to estimate face-to-face proximity [132]. Although it is not fully accurate because of the physical position of the smartphone and surrounding environments, it can still provide a trend that people have less face-to-face interactions. Hence, owing to the social distancing policy, a reduction in the number of unique Bluetooth signals is expected. Theoretically, this would indicate whether our participants adhere to the rules of staying at home and avoiding others visiting their household.

Moreover, social distancing has also affected people when they go for essential shopping. Many grocery stores have a limited number of people in their branches and have introduced directional floor markings to help shoppers maintain a 2-meter distance from one another [78]. This policy could reduce the capacity of crowded grocery stores, and fewer people are expected to be in close proximity to our participants compared to the time before social distancing measures were implemented. Thus, from Bluetooth signals, we could expect a reduction in the number of unique devices from a single scan.

4.2.4 Experiments

We report results from 2 longitudinal studies conducted to gather smartphones' digital phenotypes. Both studies were underway prior to, and continued through, large-scale transmission of COVID-19 and associated social distancing behaviors.

4.2.5 Participants

The studies were reviewed and approved by the Department of Computer Science Ethics Committee at the University. A total of 16 participants were recruited (4 males and 4 females per experiment) through the university database and websites. The 2 experiments recruited individuals from different populations in the United Kingdom: (1) students and (2) patients with a diagnosis of Parkinson disease (aged 63-75 years). The 2 studies used smartphones to capture data on the participants' activities. Both experiments rely on the same sensing platform.

4.2.6 Instrument

In this study, we used smartphones as independent sensing tools to retrieve participants' behavioral data. The AWARE sensing platform [67] and developed plug-ins were deployed on participants' smartphone as a monitor app. Under the approval of the ethics committee, different kinds of data, including calls, messages, social media app usage, smartphone usage, notifications, locations, Bluetooth signals, and Wi-Fi signals were collected passively. The content of sensitive communications, such as calls, messages, and conversations, was not recorded. All these data were processed to maintain the anonymity and confidentiality of all participants. All data sources are summarized in Table 4.1.

Participants were asked to attend an introductory interview to obtain information on our study and to clarify any of their doubts. On obtaining formal approval from the participants, the AWARE app was installed on their smartphones. Participants were asked to keep the installed app running and use their phones as they normally do. An offline analysis was conducted on data synced with the backend AWARE server.

Table 4.1: Sources of the collected digital phenotypes with data description

Source	***Data***
GPS	Location coordinates (longitude and latitude).
Weather	Temperature, humidity, pressure, wind speed, cloudiness, amount of rain and snow, times of sunrise and sunset.
Applications	App usage data.
Notifications	All notifications generated by any app installed on the phone .
Screen	Screen interactions, visited websites.
Wifi	Access points.
Bluetooth	Nearby devices.
Battery	Charging status, charging start time, charging end time, discharging start time, discharging end time.
Calls	Calls types (outgoing, incoming and missed calls) and times. Numbers are stored in an encrypted format.
Keyboard	Time and the typed letters. (numbers and emails are replaced with asterisks).
Consent form	Name and signature.
Questionnaire	We ask about interests that we infer from the data.
Activity recognition	Tilting, running, on vehicle, walking, on bicycle, on foot.

4.3 Results

4.3.1 Results Overview

This section discusses the results obtained from the responses to the stay-at-home and social distancing policies. To show how digital phenotyping can help understand behavioral responses to these policies, we selected a prototypical participant who exemplified the general behavioral responses exhibited by all participants, in each subsection, except for Figure 4.1, which represents all participants. Our experiments started at different times; therefore, the lockdown timelines for each participant may differ. The behavioral responses to COVID-19 were captured despite the differences in the lockdown week. It was intended per our experimental design to have participants adhere to these policies at different times because participants were individually assessed, and no extrapolation among other participants was intended.

Figure 4.1: Impact of the stay-at-home policy on mobility behavior.

4.3.2 Stay-at-Home Measures

Mobility patterns for participants in both experiments significantly decreased as a result of the compliance with the stay-at-home policy (P¡.001) (Figure 4.1A). Before the lockdown, the average number of places visited was slightly lesser among patients with Parkinson disease than among the students (Figure 4.1B). However, a patient with

Parkinson disease and a student may exhibit similar responses to the stay-at-home policy. Thus, individuals of the same group can exhibit a pattern that is different from the average behavior of their corresponding groups. Thus, individual analysis of digital phenotypes would help better understand people's compliance with the suggested policies.

Participants exhibited similar behavioral responses to COVID-19 regulations. We selected a participant who exemplifies the behavioral responses to present the results. We divided the participant's behavior window by week (Monday to Sunday), such that a whole cycle of a weekly social routine could be acquired. The stop point detection algorithms were applied for raw GPS data, such that the place of residence of the participant could be extracted. We used the algorithm proposed by Li et al [123] to extract stop points. The algorithm processes data points sequentially, and stop points are defined on the basis of predefined time and distance thresholds. Furthermore, we considered the location where participants spend most of their time of the day as their home. We used Foursquare [72] to determine the names of places, which allows for a better understanding of location semantics. By summing up the calculated results of the algorithm, the length of time participants spend at home and time spent by participants outside of home per week were obtained.

Another indicator is Bluetooth signals. As mentioned before, a scanned unique Bluetooth device could represent a person in close proximity. With everyone staying at home, fewer new identified Bluetooth entries were expected to be recorded. The time spent outside of home was usually below 30 minutes, but identified Bluetooth entries were all above 1000. To easily observe the similar trend of time spent outside of home and the number of new identified Bluetooth entries, we normalized the actual data so they can be plotted on the same graph. As illustrated in Figure 4.2, a clear boundary was observed, in that the participant went outside of home fewer times and presented decreased unique Bluetooth entries. Although fluctuations continue, the edge appeared around week 9; that is, March 15-22. This was the week before a lockdown was officially declared in the United Kingdom. Thus, it was observed that this participant perceived the stay-at-home policy and obeyed it objectively.

Figure 4.3 shows the impact of the "stay-at-home" policy on participant mobility. The figure represents the mobility behavior of participants who reside in the United Kingdom. Starting from week 12, the number of visited locations drastically decreased from an average of 7 locations to 2 locations. The 2 locations are the participant's home and a grocery store. To motivate this participant to comply with the stay-at-home policy, options for the delivery of grocery items or shopping times can be communicated.

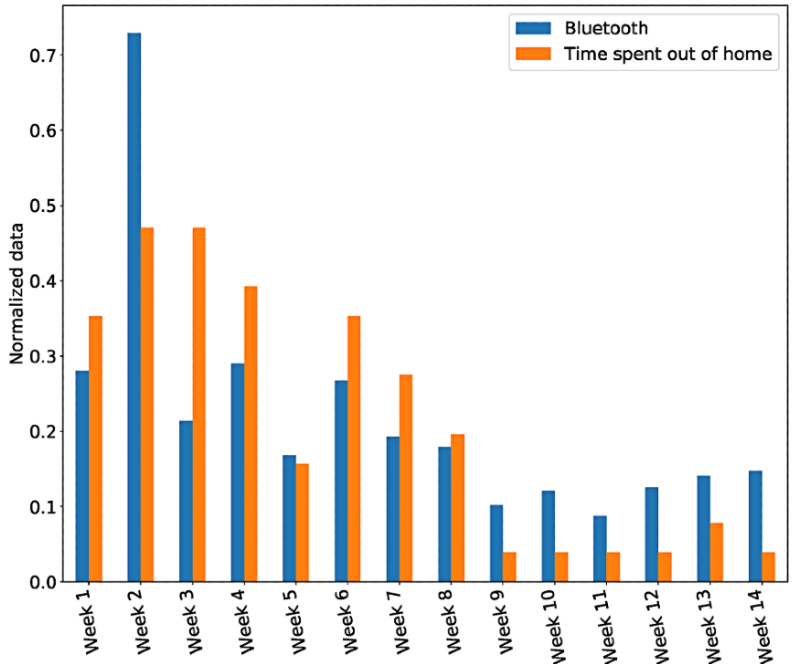


Figure 4.2: Impact of the stay-at-home policy on mobility behavior.

4.3.3 Social Distancing Measures

As described before, in accordance with the social distancing policy, people have to stay further away from each other than they would during normal times. Because of the capability of the Bluetooth technology, fewer scanned entries would be expected at a time. In this example, we also separated the data into natural weeks and combined all Bluetooth records within that week. Then, we divided this number by the total times for the scans to calculate the average 1-time Bluetooth discovery. As shown in Figure 4, the average 1-time Bluetooth entries decreased around week 9, which is the first week of the official lockdown in the United Kingdom. This potentially indicates that the participant maintained social distance with others and met fewer people during the lockdown.

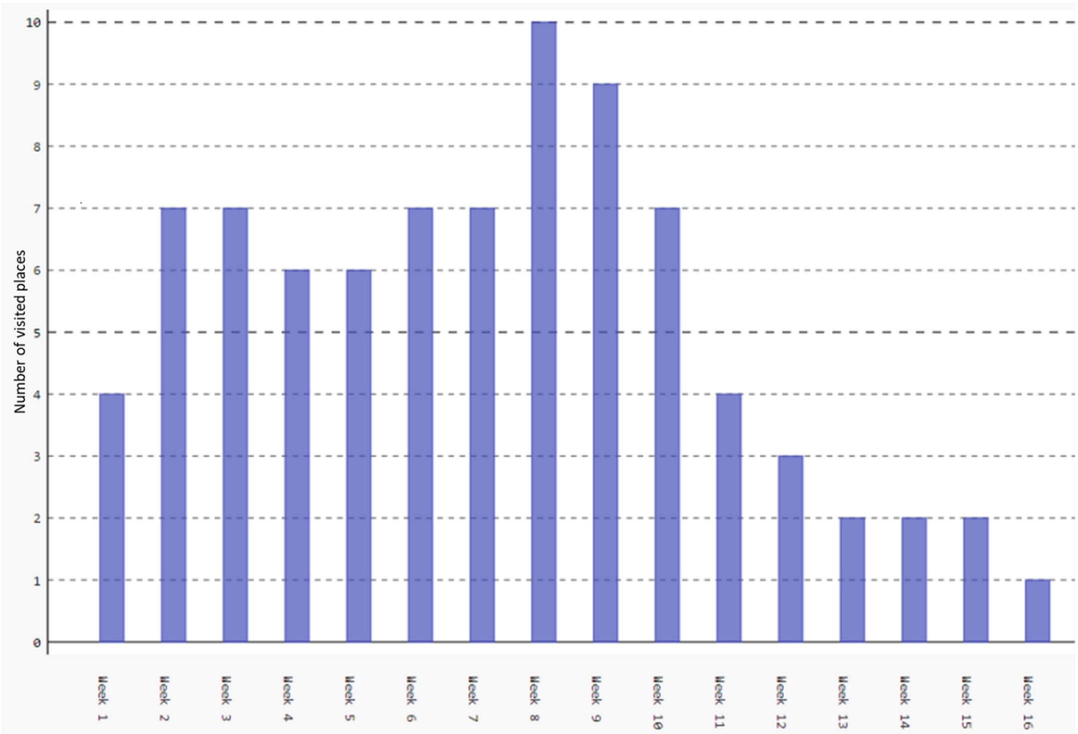


Figure 4.3: Location visited by a participant before and after the lockdown.

The results of our experiment show that the participants complied with COVID-19 policies. Participants managed to stay at home and adapt to the requested changes. However, to stay connected, the participant data show corresponding changes in app usage. The usage of social media apps, phone calls, and video conferences increased for most participants compared to the period before the lockdown. Figure 5 shows the app usage of a participant before and during the pandemic. Instagram was used the longest at 19.50 hours of usage, whereas the time spent on the Houseparty app was 9.27 hours. Values were normalized to easily observe the trend and be consistent with observations from other sources. The lockdown started during week 3. Consequently, the usage of apps, such as Facebook Messenger, WhatsApp, and Discord, has increased.

In contrast, 2 participants presented a decline in phone usage during the lockdown. When interviewed, the participants indicated that they started to use their personal computers and smart televisions more to accomplish the same tasks they previously did with smartphones.

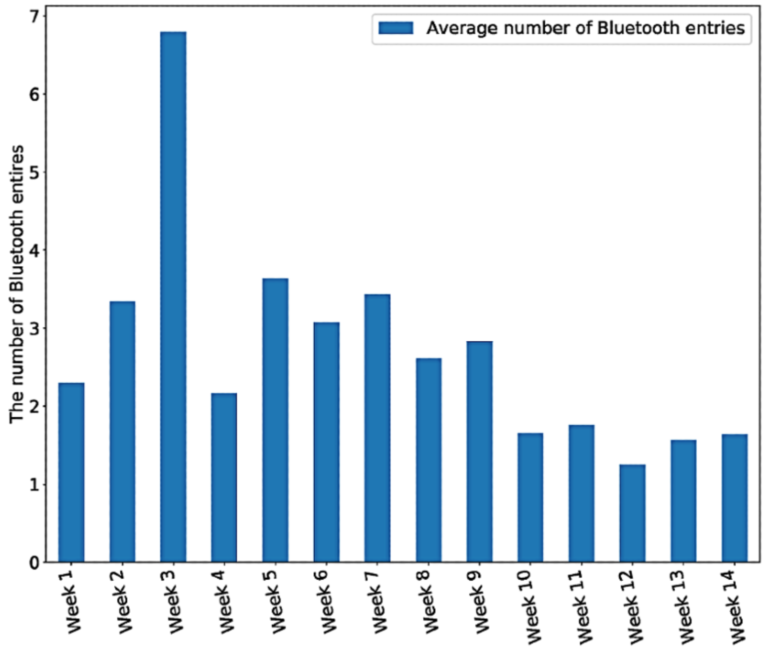


Figure 4.4: Average 1-time Bluetooth entries before and after the lockdown.

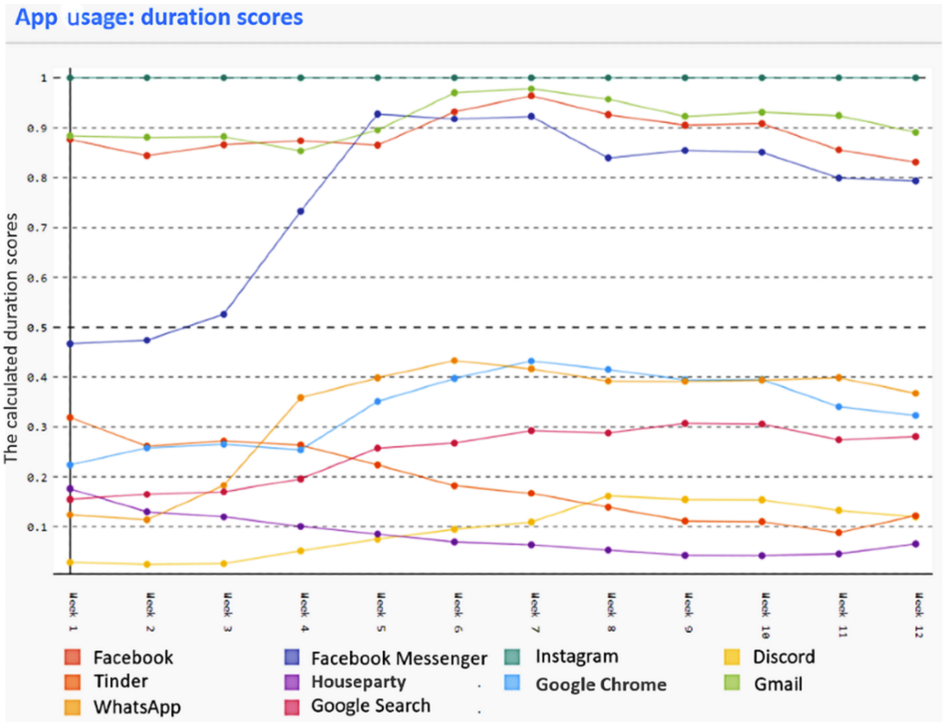


Figure 4.5: Normalized duration scores for a participant before and during the pandemic. The participant was enrolled in the first week of March and the lockdown started after the third week of data collection.

4.4 Discussion

The reported results show that actionable information can be derived from digital phenotyping. The information derived from understanding participants' compliance, as well as the behavioral impact, can be used in personalized behavioral interventions. Behavioral nudges are used as an effective approach to promote behavioral changes. The NHS in the United Kingdom employs behavioral principles, such as reducing the cognitive load, to communicate nudges. We use actual text messages delivered by the NHS during the pandemic to demonstrate the potential benefit of personalization based on digital phenotyping. We show how a personalized understanding can be leveraged for more traction nudges and just-in-time intervention. The Behavioural Insights (BI) team [223] and the NHS have collaborated to nudge approximately 2 million people through text messages. The recipients of these nudges include people at the highest risk of developing critical complications should they contract the disease. The BI team employ the following behavioral principles to produce the content of a nudge (ie, the delivered text message).

- Selection of the appropriate communication channel: since smartphone apps introduce multiple communication channels (eg, SMS, WhatsApp, and Messenger), personal preferences vary. The NHS and BI team have selected SMS as their preferred method on the basis of a study that shows that 85% of 600 participants do not mind receiving text messages on their personal devices from the NHS [224].

- Signifying the key points: owing to the limitation of text messages, the NHS and BI team have to summarize extended guidelines into short messages. Accordingly, they designed messages such that the key ideas are prioritized.

- Minimization of confusion and the cognitive load: the key ideas should be delivered in a language that is understandable by laypeople. Additionally, the messages should be clear to avoid confusion and misunderstanding that may quickly spread and negatively impact people.

- Drawing on scientific behavioral findings: insights derived from behavioral and psychological studies are used to design nudges. For instance, it has been suggested that providing the rationale can help manage people's mental health when quarantined. Accordingly, the NHS and BI team comply with that when designing nudges.

These behavioral principles are population-based, which has been reflected on the content of the nudge. M1, M2, and M3 (Table 4.2) are examples of 3 nudges that are delivered in accordance with these principles. We hypothesize that digital phenotyping can better improve the content and delivery of these nudges through personalization. For instance, the predicate of M1 can be tailored in accordance with the participant's status as follows. We can predict whether or not a person lives alone from the digital phenotypes. Accordingly, 2 versions of the message can be prepared to deliver a personalized nudge. Versions can be tailored on the basis of the predicted status, age, or other demographics predictable through digital phenotyping.

Table 4.2: Text messages used by the National Health Service of the United Kingdom for nudging and our proposed personalization.

Code	***Goal***	***NHS text***	***Personalisation suggestions***
M1	Nudge to establish social responsibility and stay connected	"If you live alone, text a friend or a family member to let them know you are following advice to stay at home until it is safer to mix with others. Plan to chat to someone over the phone at least once a day."	If a participant chats regularly or lives with others, do not send the message and prevent overmessaging.
M2	Nudge to maintain a normal routine and ease anxiety	"Try to stick as closely as you can to your typical daily routine."	If a participant frequents the cinemas, send the following message: "Watch a movie and try to stick as closely as you can to your typical daily routine."
M3	Nudge to preserve mental health	"Are there things you enjoy doing at home that you usually don't have time for?"	If a participant reports home activity, do not send the message and prevent overmessaging.

Digital phenotypes can also improve M2. For instance, an individual used to go to the cinema on Saturdays. Instead of delivering a general nudge about adhering to the typical routine, we can nudge the participant to watch a movie every Saturday during the pandemic. Thus, the typical routine can be embraced, and the delivery of the nudge can be contextualized (ie, just-in-time intervention). Adhering to typical routines can improve the mental health of individuals and reduce the negative impact of COVID-19 policies.

The information derived from digital phenotyping can also be used to prevent over-messaging. M1 encourages participants to chat with others to stay connected. If the derived data show that a participant regularly chats with others, there is no need to send M1. We speculate that crafting messages on the basis of both data and behavioral principles as well as introducing fewer messages is expected to provide better results. However, actual field testing is required to scientifically measure the real effect of doing so.

Although our approach demonstrates a potential way of producing personalized nudges, it can be reflected in existing behavioral change frameworks such as the behavioral change wheel [144]. For instance, the framework of the behavioral change wheel identifies 3 main stages to the behavioral change: (1) understanding the behavior to be changed, (2) deciding on the intervention function, and (3) selecting the mode of delivery. We profile and understand the individuals' behaviors through digital phenotyping. Incentivization and persuasion are intervention functions that shape nudging [132]. Communication as a delivery mode is then used to deliver text messages that nudge people to exhibit the desired behavior.

We are aware of the privacy concerns that may hinder the measurements and implementation of personalized nudges. However, apps can be designed in a way that allows people to partially share information in accordance with their needs. For instance, an individual may choose to share the location data only if diagnosed with COVID-19, to trace and limit the spread of the disease to others. Another individual may choose to share his/her data to receive personalized nudges that help him/her adhere to the daily routine (M2). Nevertheless, in these cases and others, personal behaviors are privately phenotyped, and it is up to the person whether or not to share the collected data. Alternatively, messages can be packaged with the app and delivered to participants on the basis of the outcome of a decision tree.

Stay-at-home, social distancing, and other policies are primarily behavioral measures aimed at changing individuals' behaviors to ensure that the risk of contracting the disease is reduced. From this standpoint, behavioral change frameworks (eg, nudging and the behavioral change wheel) can be relied upon to support the implementation of these behavioral policies. The use of digital phenotyping in activating these frameworks provides an opportunity to personalize the delivery of these policies on the basis of each individual's data. Individuals, institutions, and governments can benefit from such personalization in containing the spread of the virus. Governments may choose to develop apps that have behavioral policies implemented as built-in messages. The delivery of these messages is designed to adapt in accordance with the exhibited behaviors. Individuals who stayed at home (according to digital phenotyping) will not receive messages encouraging them to do so. This decision and others related to message delivery are made locally, on the individual's phone, without compromising his/her privacy. However, individuals who test positive can help governments reduce the potential impacts on others by voluntarily sharing their latest mobility behaviors.

Besides generating personalized nudges, digital phenotyping shows its capability to observe people's behavior on an individual level. In the contest of the COVID-19 pandemic, digital phenotyping has great potential for various implementations. Some of the COVID-19 tracking apps such as TraceTogether in Singapore and COVIDSafe in Australia have used Bluetooth technology embedded in smartphones as their primary contact tracing tool [41]. People are encouraged to install these apps so they can know if they have been in close contact with individuals who have tested positive for COVID-19. Institutions such as universities can implement digital phenotyping as innovative methods to study the traditional physiological or societal questions, since no face-to-face settlement is needed. Care facilities could also have digital phenotyping apps installed on their clients' smartphones, such that their issues can be noted without face-to-face reporting. Moreover, the large amount of personal and longitudinal digital phenotyping data could provide policymakers with a deeper understanding of the impact of COVID-19 on a sample of the population. This would shed light on how people actually react to these policies, rather than only determining the infection rate.

4.5 Conclusions

This study shows how digital phenotyping can be of value in understanding people's behavior during a pandemic. Behavioral features extracted from digital phenotypes represent the cornerstone that facilitates the personalization of and compliance with behavioral policies. We presented examples of using Bluetooth, GPS, and app usage data to analyze behavioral responses to COVID-19 policies. Additional sources can be further investigated, such as accelerometers and their role in understanding if people pause more to maintain safe distance.

To encourage the large-scale adaptation of digital phenotyping, governments can emphasize the potential benefits of public health and of maintaining mental health. To preserve privacy, an individual's data are stored locally, and he/she can make the ultimate decision on what to share and to whom the access is granted.

A rule-based messaging implementation can be used to deliver nudges on the basis of the analysis of digital phenotyping. In future studies, we intend to examine the impact of these suggested messages on a sample of the population to measure the impact of preventing overmessaging. Conducting a real-world experiment would also enable us to assess whether having more tailored messages would yield the expected benefits.

Chapter 5

Monitoring Social Withdrawal and the Impact of COVID-19

The previous chapter successfully reflects participants' conformance to the government's restrictions for reducing the transmission of COVID-19. This chapter turns to the original planning of our year-long longitudinal study for observing social withdrawal in PD patients. A series of validation measures, including Parkinson's-related psychological and clinical questionnaires and a specially designed diary, were provided to participants apart from continuous smartphone monitoring. Preliminary results on successfully tracking participants' social behaviour are explained. In this chapter, we also inspect participants' personal social changes to learn the impact of COVID-19 at an individual level. Some particular phenomena, including unusual calls and messages, were discovered in certain participants. We also had semi-constructed interviews with all participants to discuss our observations.

The content of this chapter is adapted from *Heng Zhang, Bijan Parsia, Ellen Poliakoff, and Simon Harper. 'Monitoring Social Withdrawal with Smartphones in People with Parkinson's Disease and the impact of COVID19'*. It's currently under view.

Author's contributions

Heng Zhang designed and conducted the research, collected the data, analysed and synthesised the findings, and wrote the paper. Ellen Poliakoff gave suggestions on study design and helped to conduct the study. Bijan Parsia and Simon Harper provided constant input throughout the study, including advice, feedback and critical revisions to the final manuscript.

Abstract

Parkinson's disease is a long-term neurodegenerative disease that progressively deteriorates the quality of life (QoL) of patients. Social wellbeing is an essential part of QoL. Therefore, social withdrawal, which is practically defined as reduced social interactions, could be a medical or psychological indicator for people with Parkinson's. Smartphones are social hubs of personal communication, and they are embedded with various sensors, so they are promising for social behaviour monitoring (also known as digital phenotyping). Notably, the smartphone sensing method could enable us to understand social behaviour changes on an individual level. To study social withdrawal in people with Parkinson's disease, we proposed a longitudinal study to observe actual Parkinson's participants' social behaviours via smartphones. A monitoring application was installed on participants' smartphones to capture social-related data from smartphone-mediated communications 24/7. A specially designed diary was also provided to participants to record their weekly social interaction extent and QoL. The COVID-19 pandemic, which significantly impacts social lives, provides a chance to examine this method more dramatically. Our smartphone sensing approach reflects changes in social interactions, and this meets our expectations based on the lockdowns, social distancing and isolation policies associated with COVID-19 in the U.K. The interviews of participants confirm that our observations successfully detected participants' personalised responses to the ongoing pandemic and individual adaptations. The preliminary result of the proposed social behaviour model also shows that smartphone features can re-establish participants' weekly social activity levels. The study shows the feasibility of the smartphone as an individualised social monitoring tool to reflect participants' social behaviours. Participants indeed have different reactions to the COVID-19 pandemic. Specialised services could be provided to them during this difficult time. Furthermore, it indicates that similar monitoring technology could be applied to promote personalised health services and maintain the social wellbeing of wider communities.

5.1 Introduction

Parkinson's disease (PD) is an incurable long-term neurodegenerative disease involving gradual loss of motor and non-motor functions. The most apparent symptoms of PD are shaking, rigidity, slow movement and tremor. Mood changes, cognitive decline, pain, sleep disturbance, apathy and autonomic dysfunction are all parts of the non-motor symptoms of PD [174]. From patients' experience of PD, all these symptoms significantly impact their social lives, which leads to social withdrawal. Loss and alteration of social identity have been found in PD patients [206]. Phenomena such as decreased social confidence, social anxiety or a sense of embarrassment also rises among them [206]. Disruptive social connectedness was also reported from interviews with PD patients. Overall, PD induces a decrease in social interactions, which is termed the social withdrawal of PD patients. In addition, PD progression is idiosyncratic [107], so every patient has a unique pattern of disease progression. Combined with different habits of social behaviour, each individual may have a distinctive path of social withdrawal.

On the other hand, recognising PD progression stages is the pre-requisite of PD treatment or management to maintain patients' wellbeing. Compared with the hourly and daily fluctuated symptoms, clinical assessments are applied only approximately every six months [76]. And these assessments only rely on PD patients' or doctors' experience and memory, usually in an aware situation, so they are possibly subjective, unreliable and biased [80] and unable to reflect the disease changes for the past period. Hence, a continuous, objective and unobtrusive monitoring is the pursuit of PD measurement.

Almost everyone carries a smartphone nowadays, and it has become the hub of communication. Various kinds of contact, such as messages, calls and video chatting, have been supported by smartphones. Embedded sensors also enable them to detect the surrounding environment, which can be utilised to identify face-to-face conversations. After an initial configuration, it can record data 24/7 without interruption unobtrusively. The smartphone has also become a novel research tool in understanding human behaviour, which is termed digital phenotyping. Previous studies have confirmed its feasibility and informativeness [18], and it is encouraging in expanding the knowledge of PD patients' lives [19]. The continuity of digital phenotyping also provides clinicians with a full behavioural picture of the past period rather than snapshots. So, the smartphone is promising for monitoring overall social interactions individually.

As discussed above, social withdrawal could be an indication of QoL and disease changes. The smartphone is capable of capturing the general social behaviour of every single person. Thus, a longitudinal study monitoring social withdrawal in PD with smartphones was planned. The experiment was not trying to confirm that PD can cause social withdrawal but to explore social behaviour in Parkinson's patients using smartphones. It was not aimed at clinically validated findings but a proof-of-concept study. The aim of the experiment was to test the feasibility of using the technology to observe the social lives of Parkinson's patients. Several participants were recruited to explore the relationship between social withdrawal and PD. Besides installing applications on participants' smartphones, clinical and psychological measurements of QoL, PD progression and factors causing social withdrawal were conducted every two months. A paper diary was also provided to each participant to record their social ratings weekly. All these instruments were treated as references for the smartphone data.

From more than seven million raw smartphone data points, features were built up across miscellaneous aspects of the participants' social lives. We did not intend to generalise our results to prove that Parkinson's can cause social withdrawal but rather to observe the Parkinson's patients' social behaviour and provide potential knowledge to Parkinson's researchers or carers. This paper illustrates the background and motivation of the longitudinal observation of Parkinson's patients. The detailed plan and measurement scales applied in the experiment are also revealed. The results show that our method reflected the social behaviour of these participants. Rather than investigating social withdrawal across the population, we focused on personal-level monitoring. Lockdowns introduced in response to the COVID-19 pandemic in the U.K. also allowed us to observe reduced social interactions in a more intense way. Interviews with participants examining our models can detect personal responses and the potential problems of these extreme phenomena. It implies that similar monitoring technology could be applied to other vulnerable populations to alleviate the pandemic's impact. Precision medication and personalised treatment could also be provided with the application of digital phenotyping.

5.2 Background

5.2.1 Parkinson's disease causes social withdrawal

PD is an incurable long-term neurological disease with an unknown cause. It is a degenerative disorder of the central nervous system causing progressive deterioration of brain function. More than 10 million people worldwide are living with this disease.

The typical PD symptoms are movement-related, including tremor, slow movement, rigidity, impaired posture and balance. Moreover, there are non-motor symptoms caused by PD, which include sleep disorders and neuropsychiatric, autonomic, gastrointestinal and sensory symptoms [34]. All these symptoms do not progress linearly but fluctuate and may depend on different factors, such as medication, sleep and stress [71]. This causes difficulties for the management and treatment of PD.

Various symptoms of PD contribute to social withdrawal from different perspectives. First are the motor symptoms, including shaking, rigidity, bradykinesia and tremor. These reduce PD patients' mobility, but social engagement requires people to go out, meet in groups and engage in social activities. If the necessary motor abilities are inadequate, it is natural that social engagement frequency will reduce [216]. Non-motor symptoms, including depression, apathy, stigma, social anxiety and cognitive impairment, could lead to decreased social interactions. PD is a classic example of a subcortical disorder where apathy is observed [173]. It has been found that cognitive, and particularly executive, dysfunction is often reported in PD, and those patients show the clinical features of apathy. Thus, apathy can cause PD patients to lose motivation and passion for social contact. This also applies to depression [213], stigma [139] and social anxiety [21], which destroy the incentives for social activities. Cognitive impairment could impact the social functions of PD patients [108]. For example, PD patients may find difficulty in following conversations and holding others' places, which results in frustration and neglect from others [146]. Then, PD patients could lose interest in engaging social activities gradually. Broadly speaking, social withdrawal is a potential indicator of symptoms deteriorating.

For incurable diseases, improving QoL is the aim of treatment. But from the study's QoL questionnaire with patients, it is believed the frequency and severity of non-motor symptoms are the most critical QoL predictor and contribute more than motor symptoms to QoL [141]. However, there is often a lack of awareness of the importance of non-motor symptoms [25]. Furthermore, social factors have an essential role in QoL [246]. Researchers have compared the QoL of PD patients with the general population and found that one of the areas the disease particularly interferes with is social functioning [194]. In general, social lives affected by non-motor symptoms could be a novel reflection of QoL, something that is seldom investigated in PD [246].

To our knowledge, social withdrawal is one of the crucial consequences of PD but has never been studied before. Understanding this behaviour is beneficial for the well-being of PD patients but also other diseases affecting social functions.

5.2.2 Social withdrawal

Literally, social withdrawal indicates social behaviour deviations from normal conditions. However, it is a complex phenomenon which can be influenced by a variety of factors. These include not only the impact of disease, but also sociodemographic features such as age, culture, neighbourhood, economic status and availability of transport [233]. Also, different people have distinctive social patterns. There is no universal standard to indicate that a certain amount of social activity decrease is social withdrawal. Therefore, all these variables need to be considered to arrive at a full picture of social withdrawal, which would be better studied at an individual level because all these features cannot be completely controlled to draw a population-level conclusion.

In addition, the difference between loneliness and social withdrawal should be clarified at the beginning. Humans are social animals. We all have a pervasive desire to form and maintain a minimum quantity of satisfying social relationships [13]. When the closeness of contacts does not meet a person's expectations, loneliness happens. It is a distressing feeling of perceived isolation that is highly subjective and emotional [111]. However, social withdrawal describes an objective phenomenon in which absence of social interactions occurs. It is measured by a lack of contact with social network members [157]. Although subjective feelings of loneliness and objective social withdrawal often correlate [33], they have different constructs. People may still feel lonely in a crowd of people. Since we are using the smartphone as an unobtrusive and objective monitoring tool, it is not applicable for measuring subjective feelings. We limited our scope to quantifiable reduced social interactions, which is social withdrawal.

In addition, under the COVID-19 pandemic, social withdrawal has a novel context. In the U.K., the government requires people to stay at home, reduce times going out and have face-to-face contact with only a limited number of people. Therefore, the chances of involuntary social activities are reduced. People cannot go to social places like pubs, parks or squares to enjoy casual social lives as they used to. Although new communication methods, such as video chatting, have arisen, they cannot fully replace face-to-face interactions [242]. Natural social lives cannot be achieved as they were before pandemic times. People have to socially withdraw to some extent because of the lockdowns introduced in response to COVID-19.

5.2.3 Social interaction measurement

As discussed above, social withdrawal is measured as reduced social interactions, so generally, social interactions are the actual variables to be assessed. By comparing the number of social interactions during different periods, social changes can be recognised. Social interaction is the foundation of social life. It is a sequence or aggregation of social behaviour or actions which requires a mutual orientation. People try to 'affect or take account of each other's subjective experiences or intentions' during social interactions [187]. A social target and reciprocal relationships are the critical elements of social interactions. And all social interactions happen through certain channels, such as face-to-face interactions or smartphones. Thus, by monitoring social activities that take place in these channels, the extent of overall social interactions can be inferred.

In terms of communication channels, these evolve through history. In ancient times, people relied only on face-to-face communication to connect. When paper was invented, information could be written down and transferred to a broader community. In recent centuries, more media were invented, and computer-mediated communication has become commonly applied. With the popularity of the smartphone, people spend substantial amounts of time on it for social interactions, and it has gradually become the hub of personal communication [50]. So, the channels of communication today can be categorised as smartphone-mediated and non-smartphone-mediated. For non-smartphone-mediated communication, face-to-face plays the central part. Gatherings of family and meetings of social groups are still essential elements of social life. For smartphone-mediated communication, the functionality of the smartphone provides a range of communication methods. Voice calls, video calls, emails, messages and social media are all supported. Users can choose their favourite methods to have different types of social interactions.

Typically, the methods of social interaction in the experiments include interviews, questionnaires, controlled voice or video recordings and expert observations [167]. However, there are innate deficits and limitations connected with these approaches. When participants do not remember previous events or experiences, recall biases are introduced in the questionnaire responses. Moreover, comparing with natural settings, participants may be aware of the controlled settings and behave differently, so the results could be different from the research targets [80]. Additionally, all these methods cannot continuously monitor for an extended period. Nevertheless, smartphones provide a novel method for measuring social interactions. Their capability to capture and infer social interactions enables the objective: longitudinal and in-the-wild measurement of social interactions.

5.3 Related Work

5.3.1 Smartphone social sensing

Using smartphones to investigate human behaviour has drawn great attention for decades. It is often termed smartphone-based digital phenotyping, which refers to 'moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices' [164]. As illustrated in the background, the smartphone is commonly used as a medium for social interaction. So, it is feasible to understand the social lives of users. This idea has been implemented in a number of studies. They investigate general human social behaviour [239], its relationships with personality [35], mental health (including depression, anxiety and stress) [119] and certain diseases (for example, schizophrenia [28]). In general, possible social interaction on smartphones was explored by these studies. First are calls and messages; these were monitored by contacts, length and status (incoming, outgoing, etc.) [195]. Some studies also included social media usage, but this cannot be captured as precisely as calls and messages due to system restrictions. Therefore, it is often inferred by the application usage [74]. It is possible to record social interactions happening outside of the smartphones as well those mediated by the phones. Since everyone usually carries a smartphone and the Bluetooth embedded on the smartphone keeps scanning surrounding signals, a scan entry could represent a person [250]. Therefore, Bluetooth is employed as an inference for people proximity. The microphone is another typical environmental sensor on smartphones that is utilised as an indicator of social interaction. Raw audio captured by the microphone was processed to detect whether conversations were happening around the person carrying the smartphone [239]. Then all these features were processed and correlated with the target of these studies. For example, a study of personality [237] explored the relationship between the collected smartphone data and the big five personality traits. It found that certain features are strongly correlated with the different traits, such as a positive correlation between Bluetooth entries and extraversion but a negative one regarding agreeableness. All these studies benefit from the unobtrusiveness and continuity of the smartphone as a novel research tool. The burden of participants is released, and both contextual and behavioural data can be monitored, allowing researchers to observe changes and deviations from a comprehensive perspective [11] [202].

5.3.2 Digital phenotyping in Parkinson's

The smartphone has also been employed to study Parkinson's patients' behaviours, but has mainly been confined to motor functions. A smartphone application that assessed voice, posture, gait, finger tapping and response time were created by [8] to classify the motor impairment of Parkinson's. [117] evaluated the tremor intensity from the accelerometer of the smartphone. The strong correlation between the result and clinical scale of Parkinson's shows the smartphone's potential in assessing the severity of the symptoms. However, these studies all required patients to do specific tasks, which were intrusive. [94] generated mobility features from longitudinal passive smartphone data to estimate Parkinson's patients' daily motor symptom fluctuations including pain, gait, freezing and fatigue. Gait abnormality was captured by [260] to monitor the medication adherence of PD patients. These studies applied passive smartphone technology, but they only focused on motor symptoms; QoL and social impact were still being neglected. To our knowledge, the social withdrawal of Parkinson's has rarely been studied by smartphone social sensing. The latest smartphone social sensing technologies are implemented to observe the social lives of Parkinson's patients.

5.4 A general model of social interaction

As described in the background, we believe there are two types of communication channels. One is smartphone mediated and another one is non-smartphone mediated, which typically means face-to-face interactions. Hence, overall social behaviour will be outlined by summarising these two channels.

Non-smartphone-mediated communication implies direct interactions, which are face-to-face. The smartphone provides various channels of communication, including calls, messages and social media usage. Face-to-face is probably the dominant mode of all interactions [15]. Gestures, expressions and voice interactions are conveyed during face-to-face interactions. With calls, the approach narrows to the verbal method only, with pitches and tones to help participants understand each other. Limitations increase in messages, which rely on text. And social media can involve even less: hitting a 'like' button could be an interaction. So, face-to-face provides the most abundant medium of communication, followed by calls, then messages and, lastly, social media. Moreover, there is evidence that interactions via voice create stronger social bonds than interactions only including text [116].

Using this knowledge, the number of social interactions in each channel was typically measured by times, length and unique contacts [257]. Times represented the frequency of social behaviours conducted in a certain time, usually a week, fortnight or month. Every call made, message sent, social media use and face-to-face interaction is regarded as one social activity. However differently, the duration of each communication carries an inherent meaning. For calls, it indicates the total length of each call. For messages, it calculates the effective part, which is the number of characters. And in face-to-face interaction, the length is measured from the beginning of the first word to the end of the last sentence. Unique contacts signifies the diversity of the contacts in this communication channel. People could make a number of phone calls to a single person, which could generate a high total length but few unique contacts. So, unique contacts provides another perspective to quantify the extent of social interaction. Face-to-face interactions were inferred from the built-in plugin of AWARE. It scans surrounding sound via microphones of smartphones. If the captured sound is inferred as the human voice, this conversation starts until this sound disappears. This period is counted as a conversation session. By combing all sessions into a specific length, like a week, the total number and length of conversations can be known. Social media usage was also calculated by session. If the social media application is in the foreground, the user session starts until this application exits. By combing all sessions, the number of times and total times people spent on social media were acknowledged.

5.5 Longitudinal one-year data collection

From all the theoretical background explained above, we planned and started an exploratory one-year longitudinal study to observe the social behaviour of PD patients.

Participants were recruited on an enrolment-and-go basis. Advertisements were through the Parkinson's UK website. We did not specifically qualify our participants because we were focusing on an individual-level observation. We only required participants to be regular smartphone users (self-described) and score higher than 88/100 on the Addenbrooke's Cognitive Examination Revised (ACE-R) scale [147], which means they have no cognitive impairment [94]. If potential candidates state that they have other diseases having more life impact than PD, they were also excluded. Once a potential candidate contacted us, we would have an interview with them to check if they met the basic standard of participant requirements. Then, the application would be installed on their smartphones and the data collection would start. This experiment had two major components: the continuous smartphone data collection and the traditional mechanism of measuring clinical, psychological and subjective input from participants as ground truth. The full data collection plan is shown in the figure 5.1.

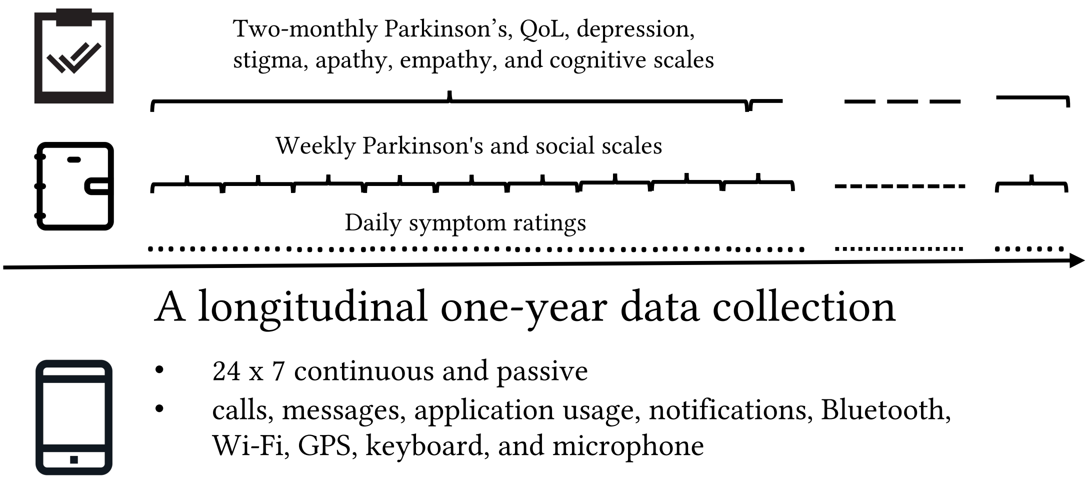


Figure 5.1: The plan for longitudinal one-year data collection.

A monitoring application was chosen to record all designated sensor data on smartphones. It is called AWARE [67] and is one of the full-scale sensing platforms, providing plenty of functionality for our needs [245]. According to the systematic review of smartphone social sensing [257], we collected data from all possible sources related to users' social interactions. These sources included calls, messages, application usage, notifications, Bluetooth, Wi-Fi, GPS, keyboard and microphone. Details of data collected, purposes and structures are illustrated in the table below 5.1. Due to the limitations of iOS, some significant data could not be captured on phones using that system.

We did not provide extra phones to participants since they may not have transferred all social interactions to the secondary phone. Therefore, we only recruited participants using Android devices. All collected smartphone data were ethically considered. Irreversible encryption was applied to contact IDs and Bluetooth/Wi-Fi addresses/IDs. For the keyboard, only the number of characters typed could be known, and all raw audio captured on smartphones was processed locally. We only recorded audio if it was a conversation at that moment, using the algorithm developed in [239].

Table 5.1: Data source, purposes and structures of collected sensor data.

Data source	Purpose	Structures
Calls	Call events	Timestamp, contact ID, length, status
Messages	Messages events	Timestamp, contact ID, status
Application usage	Time spent on social media	Open timestamp, app name, app package name
Notifications	Estimations of social media messages numbers	Timestamp, target application name
Bluetooth	Estimations of face-to-face encounters	Timestamp, Bluetooth address, Bluetooth ID
Wi-Fi	Estimations of locations	Timestamp, Wi-Fi address, Wi-Fi ID
GPS	Locations	Timestamp, longitude, latitude
Keyboard	Estimations of social media messages length	Timestamp, app name, app package name, length
Microphone	Detection of surrounding sound	Timestamp, is conversation

We also applied various widely used instruments to gather standard PD progression, QoL and related clinical and psychological factors. The severity of PD was measured by a standard clinical scale, the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [76]. The famous Parkinson's Disease Questionnaire (PDQ-39) was also employed [102]. From the literature illustrated in the background, depression, stigma, apathy, empathy and cognitive impairment are probably the most direct causes of social withdrawal introduced by PD. Hence, the scales of these factors were included. Moreover, a modified social withdrawal scale from motor neurone disease was implemented to measure social withdrawal [185]. These questionnaires were conducted every two months. Details of questionnaires applied and their purposes are shown in the table 5.2.

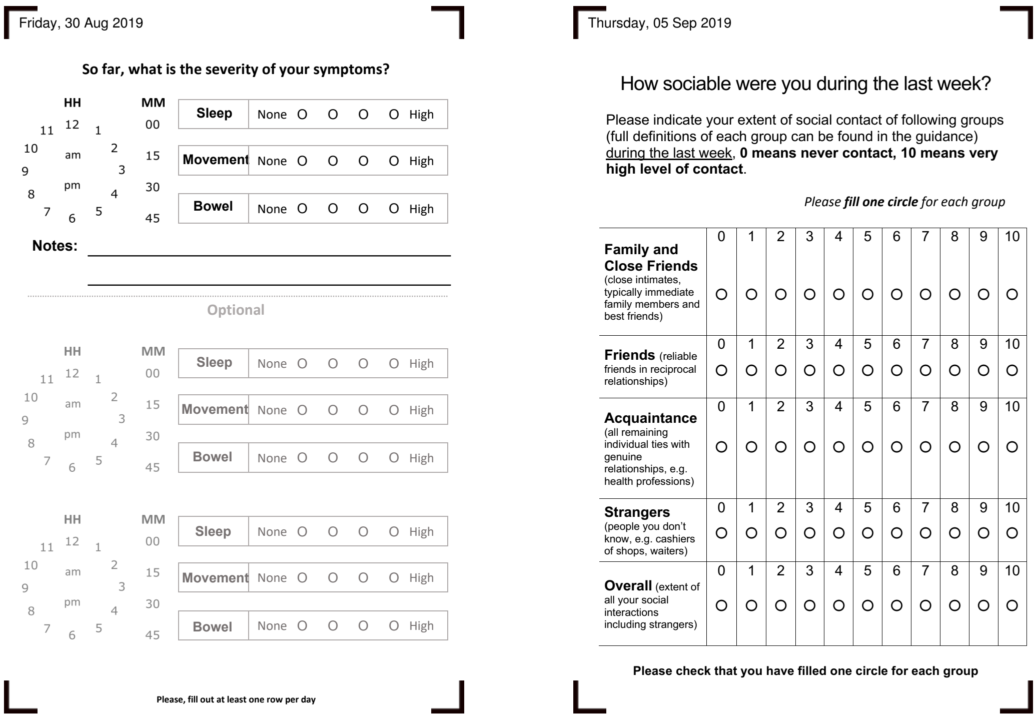
Table 5.2: Scales applied in the study

Scale name	Purpose	Reference
MDS-UPDRS	Clinical Parkinson's progression	[76]
PDQ-39	QoL impact by Parkinson's	[102]
The Addenbrooke's cognitive examination revised (ACE-R)	Cognition	[147]
Stigma scale for chronic illness 8-item version (SSCI-8)	Stigma	[151]
Geriatric depression scale (GDS)	Depression	[252]
Interpersonal reactivity index (IRI)	Empathy	[46]
Apathy scale (AS)	Apathy	[211]
Social withdrawal scale (SWS) modified (see Appendix B)	Social withdrawal	[185]

A diary was also provided for participants to record their daily symptoms, weekly QoL and social interaction scale. This diary was originally designed by [234] to track PD day-to-day fluctuations. They attempted four prototypes using Bluetooth, NFC and a microcontroller but accomplished higher acceptance and compliance using a paper diary. Participants choose three main symptoms to record, which are unique and personal. Since we are exploring PD-caused social withdrawal, a weekly social scale for confirming the interaction extent of different types of contacts was added to the diary. The participants also rate their overall social interaction level from 0 to 10. We followed the design implications in [234] and used the same circle to maintain consistency. An eight-item version of Parkinson's' disease questionnaire PDQ-8 [103] is also included in the diary each week to produce fine-grained disease progression ratings. An example of the daily symptom diary (left) and weekly social interaction scale (right) is shown in the figure 5.2.

5.6 Social fluctuations monitoring

Participant recruitment started after all the procedures of the experiment had been ethically approved. We did not require all participants to start the data collection at the same time. Once the participant had enrolled, the data collection began. At the first session, the diary was given to the participant, and the first round of questionnaires were conducted. Thus, there were various time lengths for each participant. Parkinson's


Friday, 30 Aug 2019

So far, what is the severity of your symptoms?

HH	MM		
12 1 2 3 4 5 6 7 8 9 10 11 am pm	00 15 30 45	**Sleep**	None O O O O High
		Movement	None O O O O High
		Bowel	None O O O O High

Notes: ______________________

Optional

HH	MM		
12 1 2 3 4 5 6 7 8 9 10 11 am pm	00 15 30 45	Sleep	None O O O O High
		Movement	None O O O O High
		Bowel	None O O O O High

HH	MM		
12 1 2 3 4 5 6 7 8 9 10 11 am pm	00 15 30 45	Sleep	None O O O O High
		Movement	None O O O O High
		Bowel	None O O O O High

Please, fill out at least one row per day

Thursday, 05 Sep 2019

How sociable were you during the last week?

Please indicate your extent of social contact of following groups (full definitions of each group can be found in the guidance) during the last week, **0 means never contact, 10 means very high level of contact**.

*Please **fill one circle** for each group*

	0	1	2	3	4	5	6	7	8	9	10
Family and Close Friends (close intimates, typically immediate family members and best friends)	O	O	O	O	O	O	O	O	O	O	O
Friends (reliable friends in reciprocal relationships)	O	O	O	O	O	O	O	O	O	O	O
Acquaintance (all remaining individual ties with genuine relationships, e.g. health professions)	O	O	O	O	O	O	O	O	O	O	O
Strangers (people you don't know, e.g. cashiers of shops, waiters)	O	O	O	O	O	O	O	O	O	O	O
Overall (extent of all your social interactions including strangers)	O	O	O	O	O	O	O	O	O	O	O

Please check that you have filled one circle for each group

Figure 5.2: An example of the daily symptom diary (left) and weekly social interaction scale (right).

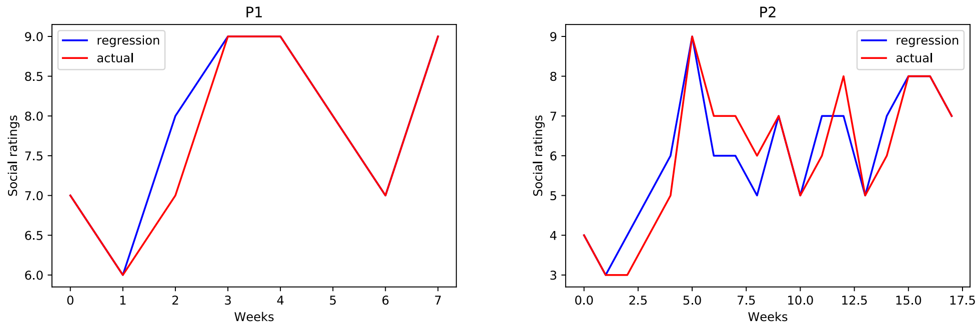
patients are typically elderly and may experience some accessibility issues. Therefore, smartphones are not as popular among PD patients as in the general population. Also, since this was a long-term longitudinal experiment rather than a one-time session, involvement the study was a long commitment for participants, and our study may have been less attractive for volunteers. For all the above reasons, we did not expect a specific number of participants to generalise our results. In addition, our study was an exploratory study, and as it was estimated that each participant would generate millions of raw data points, we aimed to involve 10 participants. The participant recruitment campaign started in October 2019 and ended in March 2020. Two participants dropped out, but the data collection continued for eight participants. The age of participants ranged from 63 to 75 years, and the participants consisted of six males and two females. They had all been diagnosed with Parkinson's for at least five years, and all their symptoms were mild, usually not interfering with their daily lives. The most common symptoms impacting their lives were sleep problems, stiffness and slowness. Only one participant had a moderate tremor, and all of them were taking prescribed medication to control their symptoms.

The COVID-19 outbreak was declared a pandemic in the U.K. in March 2020, and national restrictions and lockdowns were introduced to limit possible face-to-face contact and enforce social distancing. These measures had a tremendous impact on people's lives, especially social behaviours. People could not move freely and have social interactions as they used to. In particular, their chances of having face-to-face interactions with people not living in their households were curtailed. People had to socially withdraw to some extent, so it was not reasonable to treat their social lives during the pandemic the same as before. Hence, we divided the data analysis into two parts: the first was pre-pandemic, when social interactions could be engaged in freely without any constraints; the second was during the pandemic, when people had to follow the rules and limit their face-to-face interactions. For the pre-pandemic period, the target was to reconstruct their social interaction levels from collected objective data. As explained in the background, all social interactions have to be conducted through a particular media, which we categorised as face-to-face, call, message and social media. So, the overall social interaction should be aggregated from all these categories. The formula is $Total\,Social\,Interaction\,Level = Face-to-face+Calls+Messages+Social\,media$. However, the ground truth we relied on was still a variety of questionnaires. This has natural deficits as a subjective measurement, where recall bias and halo effect could occur. When participants did their ratings, they probably had different weights in their opinions for each communication channel. For example, they might overrate face-to-face interactions but underestimate message communications because face-to-face conversations make a bigger impression on them. Besides, the number of social interactions in each channel is determined by three typical factors: the number of unique contacts, contact length and contact times [60]. However, each individual might have particular preferences for estimating social interaction levels from these factors. For example, people having the same length of calls may treat them as a different number of social interactions because these calls could be made to a single person or to several people. Thus, a method was necessary to differentiate which factor played a more critical role when judging the extent of social interaction. For example, the calls could be weighted by this formula: $Social\,interaction\,level\,of\,Calls = \alpha_1 * unique\,contacts + \alpha_2 * contact\,times + \alpha_3 * contact\,length$. And all parameters could be zero.

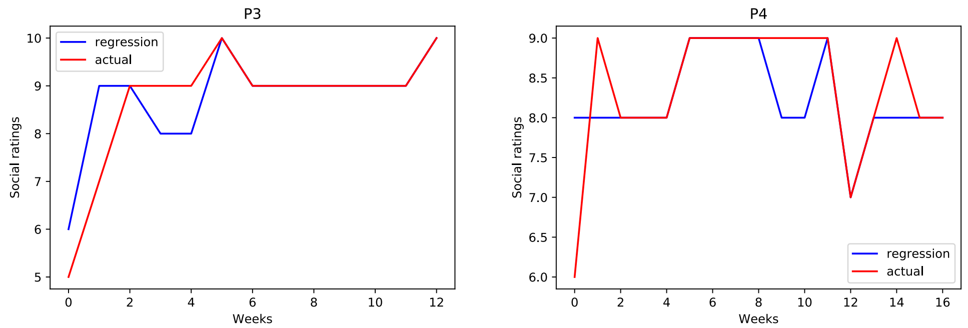
Linear regression is an appropriate fit for the target variable. It can model multiple explanatory variables to a scalar by assigning different parameters. A relationship will be found between these independent variables and the target variable by minimising the distance between observations and expected values. By integrating the number of social interactions in different channels and accrediting them with reasonable weight, we can establish the relationship between participants' weekly social ratings and smartphone data. Since all the given data are expressly limited to this person, this function will be overfitted to the specific individual, so a personalised model is generated.

All participants who had data collected in the pre-pandemic period were processed by linear regression. All three factors of each communication channel were extracted as features. Due to the limitations of the system, these factors could not be captured on social media or message apps, such as WhatsApp and Skype. Therefore, we used the notifications of these applications and number of characters typed in these applications as the estimation for contact length and contact time. Moreover, face-to-face interaction times, length and unique contacts were inferred by conversation detection and Bluetooth. Location features, including time spent at home and travel distance, were also added as additional elements because locations outside the home, such as urban public places, provide arenas for social opportunities.

The whole dataset was split into a training part to establish the linear regression model and a test piece to test the performance. All possible combinations of generated features were attempted to identify the best mixture. We used R-squared, which describes the proportion of the target variable's variance explained by independent variables, to choose the best model. This is commonly used in regression analysis, and closer to 1 means a better fit. Since all data were time-related, and the future features could only be predicted from the past, we only split these entries according to time. Features generated from past dates were always treated as training sets, and they were later used to predict the diary entries. This is an ongoing project, so not all participants have collected enough data to demonstrate the final results. We present four participants as examples to exhibit our principal models and methods. The results of these participants are shown in figure 5.3. Two other performance indexes were also calculated to examine the performance of the model. These are root mean squared error (RMSE) and mean absolute error (MAE), which measure the magnitude of the error. As described in the figure, our models achieve fairly good fits for the ratings in general. The RMSE and MAE are no more than 0.5. For particular participants, such as P3 and P4,



RMSE = 0.353553 MAE =0.125 R-squared = 0.894737 RMSE = 0.707107 MAE =0.5 R-squared = 0.833333



RMSE = 0.733799 MAE =0.384615 R-squared = 0.662963 RMSE = 0.685994 MAE =0.352941 R-squared = 0.326733

Figure 5.3: Linear regression model performance on each participant.

R-squared values are less than those for P1 and P2, but the value is still positive. This is because there are limited variations in their social ratings. For P3, nine was answered eight times in the weekly social ratings for 13 weeks in total. A similar situation occurred with P4: nine was answered nine times in the weekly social ratings out of 17 weeks. Thus, there was not enough training for the model to differentiate the ratings.

5.7 The impact of COVID-19

The COVID-19 pandemic began in March 2020 in the U.K., and the entire nation started lockdown on March 23rd, together with enforced policies to reduce the spread of COVID-19. These measures included staying at home, which required people to remain home as much as possible and only go out for essential activities, and social distancing, which limited face-to-face contact to within the household, with physical distancing required when with others. All these policies severely impacted people's social lives. People's opportunities to have face-to-face interactions were extremely restricted. They could not go out for routine activities like shopping and dining or have

Table 5.3: Social interaction level changes before and during the COVID-19 pandemic.

		P1	P2	P3	P4
Face-to-face	Times	↓	↓	↓	↓
	Length	↓	↓	↓	↓
	Unique contact	↓	↓	↓	↓
Calls	Times	↓	↓	≈	↓
	Length	≈	↓	≈	↓
	Unique contact	↓	↓	≈	↓
Messages	Times	↓	≈	↑	↓
	Length	↓	N/A	↓	↓
	Unique contact	↓	N/A	≈	↓
Social Media	Times	↑	N/A	N/A	↑

unplanned social interactions. Although we could not monitor our participants in normal conditions, it gave us an opportunity to investigate how our participants adapted to the impact of these policies. In addition, from the literature, Parkinson's-induced social withdrawal could be slight or insignificant. It could be hard to detect these minor changes. But the enforced lockdowns reduced the opportunities for social activities significantly. It exaggerated the situation of social withdrawal and applied to every individual. This provided us with a unique opportunity to assess whether our monitoring methods could detect these dramatic changes. The three factors (i.e., number of unique contacts, contact length and contact times) for each communication channel were gathered and compared with the pre-pandemic period, if possible. The results are shown in table 5.3. We used the same length of time to compare the changes for every feature in the table. Due to technical issues, certain types of data could not be collected for comparison, which are marked N/A in the table. As can be read from the table, most features decreased. But some features in one channel changed asynchronously, which are marked in black boxes. This phenomenon draws special attention and will be discussed below.

Times spent outside the home are also analysed and displayed in the figure 5.4. All the raw GPS points were put through a stop-point detection algorithm to cluster the places where the participants stayed more than 10 minutes. This calculates the distance of the stop point from the home and the difference between the current position and the previous one. So, if the participant left home for other places, the algorithm would

notice that. The whole cycle of leaving home and coming back was counted as left-home time. As indicated from previous literature, we predefined 10 minutes and 100 metres as the threshold to define a new place [204] [24]. Thus, when all GPS points are within the range of 100 metres within more than 10 minutes, the algorithm regards the participant as having visited this place.
As illustrated, the left-home times significantly decreased after the official lockdown date. Participants stayed at home as much as possible and only went out one or two times per week after lockdown. One participant never left his/her home. Besides, we also applied Foursquare, the service for semantic places to explore what kinds of places participants went. The results show they went out for only two purposes: walks and groceries. None of the outside activities were for social interactions.

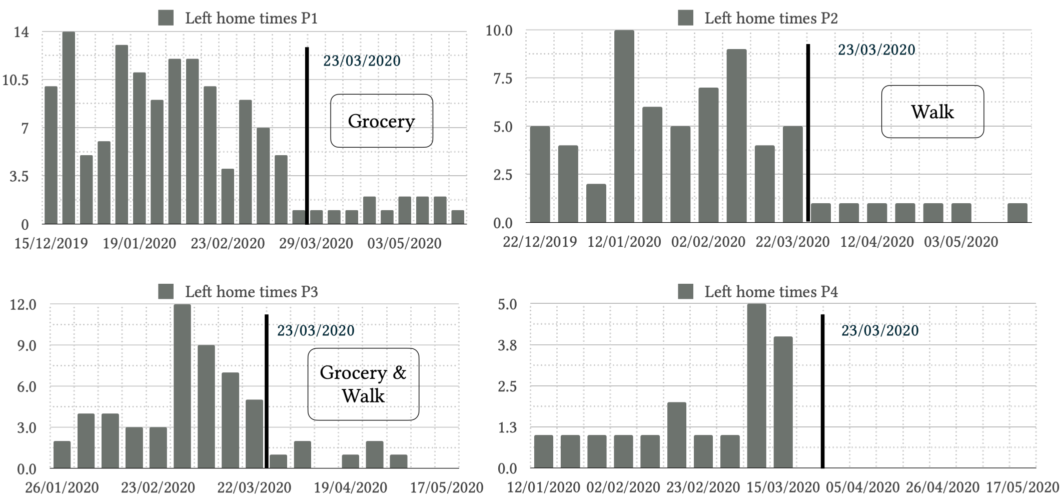


Figure 5.4: Left home times before and during the COVID-19 pandemic.

As for social interactions, there was a trend showing that all three factors declined or remained similar in the same way. Two participants showed increased usage of social media, and two exhibited unique change patterns for calls and messages. The contact times and unique contacts of calls dropped in P1, but the length remained similar. So it seems that the participant had more long calls during the pandemic (as shown in the figure 5.5). In other words, the calling pattern changed from short and frequent calls to long and infrequent calls. This was probably because short calls were used to arrange face-to-face meetings pre-pandemic. In the new situation, the participant could only make long calls, which could have been a substitution for face-to-face interactions. Similar circumstances were also observed with P2. Although the length of messages declined for P3, the unique contacts remained stable, and the times even increased (as shown in the figure 5.6). Therefore, P3 tended to send short, frequent messages during the pandemic rather than long, infrequent messages pre-pandemic. Was that because reduced face-to-face interactions made P3 need more frequent contact? To confirm our findings and investigate these special phenomena, we conducted a semi-structured interview with participants. In general, we generated questions from changes in these factors and asked participants if they agreed with them. For example, if call times decreased, we would ask the participant if he/she agreed with the data analysis that he/she made fewer calls. Moreover, their social routine adjustments and novel methods of communication were also explored in the interviews. Participants were asked about their feelings regarding their social lives under the lockdown and about the effectiveness of video calls.

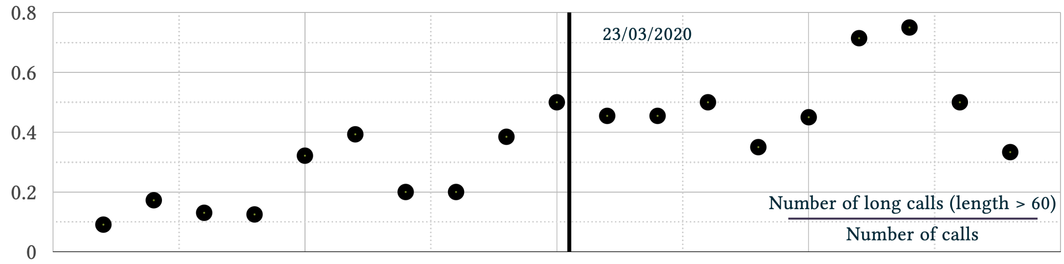


Figure 5.5: Percentage of P1 long calls before and during the COVID-19 pandemic.

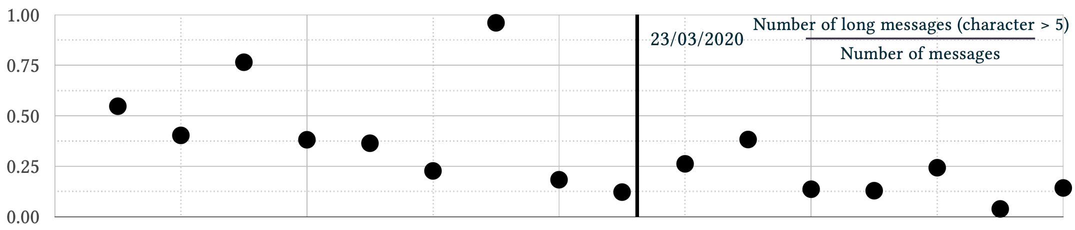


Figure 5.6: Percentage of P3 long messages before and during the COVID-19 pandemic.

All participants agreed with our observations about the trend of their communications. They admitted that all kinds of social interaction had decreased since the lockdown.

One participant, however, denied that social media usage had increased. The participant thought the time spent on social media was similar to before the pandemic. *'I don't think I use more social media than I used to, at least I am not aware of it.'* she said. For two particular participants, P1 stated, *'I made more long calls with friends and families I was able to meet often.'* But P1 did not consider long calls as an alternative to face-to-face conversations. *'There is no choice,'* P1 said. As for P3, the participant gave us a possible reason, and it was not him who needed more attention. After the start of the pandemic, P3's family was concerned about his health condition. *'They set up a family group chat, so I always check in there,'* said P3. So, he sent more *'fine'* messages in the group chat, which resulted in more short, frequent messages.

In general, all of our participants made significantly more video calls compared with pre-pandemic times. This included participants who had never made video calls before. However, they agreed that their social activities inevitably declined, and none of them deliberately increased their use of any communication channels to compensate for the lost social time. Moreover, none of the communication channels substituted for face-to-face interactions, so diminished face-to-face interactions were not replaced. P1 also emphasised that *'video calls can never replace face-to-face interactions.'* Mainly, they all missed the times when people could meet and talk freely. The interview also confirmed the participants' reasons for leaving the home. They went out either for walks or groceries. Participants who had children living around them only went out for walks because their children could deliver groceries for them. This benefited this vulnerable population, so they did not have to go to crowded places, such as supermarkets.

5.8 Discussion

The results show that these features can reflect the social interaction levels of PD patients. Smartphone monitoring is useful for understanding the social behaviour of this population. This technology contributes to the knowledge for Parkinson's patients and carers in various ways. Neurologists and Parkinson's nurses can understand their patients more granularly rather than relying only on the six-monthly check-up. Patients' social interaction fluctuations will be shown in a more precise way, and the detected social withdrawal will give carers a signal that a particular PD patient needs special attention. Thus, better treatment could be provided promptly and properly. Besides, PD patients' caregivers can also acknowledge the QoL of PD patients from a social perspective. Some situations that patients feel awkward to reveal could be reflected, so people caring for PD patients can give personalised social support for patients to maintain their social wellbeing.

The reduced social interactions due to the COVID-19 pandemic across all communication channels are reflected in the feature data we collected. The agreement from participants with our observations indicates that our methods can reflect these significant changes. From the behaviour shift following the arrival of COVID-19, all of our participants obeyed the government's policies. They stayed at home and minimised their opportunities to leave the home but tackled this situation differently. Participants who did not have family nearby had to do their own grocery shopping, Location data from smartphones can show which participants do essential shopping themselves. Quiet times to visit the shops or deliveries could be suggested from the smartphone data to reduce the chances of unnecessary face-to-face contact.

Negatively, the impact on participants' social lives was not compensated for. None of the existing channels of communication increased. Although all the participants started using video calls, these did not fulfil the same role as face-to-face interactions. Their symptoms had already caused potential social withdrawal. These severe restrictions only exacerbated the situation, reducing the opportunity for social activities. Special consideration is necessary for PD patients to maintain their social wellbeing. Although it could be different from the participants' viewpoint, smartphone data can suggest abnormal emotional states, and extra attention or interventions could be offered to these individuals. Unusual signals and pattern changes, such as P1's long calls and P3's messages, are typical examples. With more expansive applications of the technology, similar signals could be captured for each person, and special care could be offered according to these signals to mitigate the impact on this vulnerable population.

5.9 Challenges and limitations

PD patients include many older adults, and they may suffer from other diseases, such as diabetes. Their motor and cognitive functions may influence their judgement and memories. Combined with natural deficits, such as recall bias in questionnaires, answers from patients cannot be fully trusted as the ground truth.

After the beginning of the COVID-19 pandemic and the resulting restrictions, people could not continue their everyday lives. Hence, we only had a limited number of pre-pandemic weeks to test the model. The model overfits the participants to achieve a personalised understanding, but inadequate data could result in unreliable observations in the long run.

Moreover, the pandemic caused declines in social activities but fostered a novel method of communication: video calls. From the interviews with our participants, none of them had video calls on their smartphones but on tablets or personal computers. This was beyond our monitoring method's range and may cause a gap between the intended observation of overall social interaction and the collected data.

For the convenience of participants, the two-monthly questionnaires were conducted at their homes. As they all lived in different communities, this made the whole interview process time-consuming. Following the pandemic, the interview transformed to online, so some physical measurements made by touch and feel could not be made. This led to the incomplete assessment of the Parkinson's progression measurement. A similar situation also occurred with the smartphone data collection. It was not perfect all the time. The application stopped working occasionally, and participants sometimes did not use their smartphones.

5.10 Future work

The practical meaning of the models needs further investigation, and these models will also be applied to the data during the pandemic to examine if they are still applicable. More features to differentiate the types of contact will be considered to build a more explicit social withdrawal estimation. Data from different instruments, such as questionnaires, daily symptoms and weekly PD questionnaires, will also be included to examine the relationship between social behaviour and PD symptoms.

COVID-19 has shaped social activities in various ways. Therefore, a calibration for the percentage of time that participants use smartphones is necessary. A questionnaire regarding how much they use different devices for communication will be proposed.

Moreover, rather than focusing on the questionnaire passively, we plan to ask the participants questions directly to confirm or reject our findings about changes in their social interaction levels. Their smartphone usage will also be queried to estimate the difference between the objective data and participants' perceptions.

5.11 Conclusion

From the theory and practical knowledge of social behaviours, we constructed a model for measuring social withdrawal. The overall interaction model has been examined with actual data, showing reasonable performance for understanding social interaction levels. The monitoring of social interactions could also provide understanding of PD from a QoL perspective and in a longitudinal way. This is valuable for both doctors and caregivers of PD patients.

Our monitoring technology was also able to demonstrate the significant changes caused by COVID-19. Participants' personal responses and special social adjustments were also illustrated from the features we created. Smartphone monitoring was particularly beneficial during this critical period. Patients' personal difficulties could be exhibited from location data. Thus, individualised support could be provided to this vulnerable population to maintain their wellbeing.

The popularity of smartphones is higher in younger generations, who will get older and may face health problems in the future. Digital phenotyping is highly applicable for personalised disease management, not only for Parkinson's but also for people with mental health disorders or hearing loss. Customised treatment plans and precision medicine could be organised based on the granular smartphone data. We are on the way to contributing to better healthcare for every human being.

Chapter 6

Social Withdrawal and Parkinson's

The previous chapter presented the plan of the year-long longitudinal study and the personal social impact caused by COVID-19. Both preliminary results of social behaviour tracking and reflections of social changes by COVID-19 exhibit the potential of our smartphone social sensing method. With the completion of the whole year of observation, the longitudinal smartphone behaviour data and corresponding scales/diaries are finally available. In this chapter, we discuss the relationship between two-monthly conducted clinical/psychological scales and smartphone data. These scales are valid indicators of clinical or psychological signals as standardised monitoring equipment, including Parkinson's progression, quality of life, social withdrawal, cognition, apathy, empathy, and stigma. The constructions of higher-level features from raw smartphone data are also illustrated in this chapter. All kinds of social behaviours and related factors are considered, including calls, messages, social media, face-to-face conversations, and locations. A conceptual social interaction model regarding both types of contact and channels of communications are established.

The content of this chapter is adapted from *Heng Zhang, Bijan Parsia, Ellen Poliakoff, and Simon Harper. 'Exploring Social Withdrawal with Smartphones in People with Parkinson's Disease: A Longitudinal Study'*. It's currently under view.

Author's contributions

Heng Zhang designed and conducted the research, collected the data, analysed and synthesised the findings, and wrote the paper. Ellen Poliakoff gave suggestions on study design and helped to conduct the study. Bijan Parsia and Simon Harper provided constant input throughout the study, including advice, feedback and critical revisions to the final manuscript.

Abstract

Parkinson's disease (PD) is a long-term neurodegenerative disease that impacts patients' quality of life (QoL). Social life is an essential part of QoL. Both symptoms of PD and deterioration of QoL could cause reduced social interactions. This phenomenon is known as social withdrawal and could be a general health indicator for people with PD. People spend significant social time on smartphones nowadays, and under the concept of digital phenotyping, smartphones are promising for social behaviour monitoring. Considering related social factors, we initiated a year-long longitudinal study using smartphones to study social withdrawal in PD. Weekly diaries and sets of scales, including PD progression, QoL, cognition, apathy, empathy and stigma, were also conducted every two months to correlate with smartphone data. After 24/7 continuous monitoring for a year, 23 features were generated from more than ten million raw data points collected from eight participants. Since numerous uncontrollable factors could influence people's social behaviour, these data were analysed individually. At least one smartphone feature for each participant was found to have a significant correlation with standardised scales. Most of the results (81%) imply that participants tend to have reduced social activities when situations such as PD progression, QoL and apathy are worse. The correlation results between social withdrawal scales and smartphone features confirm that our approach can reflect participants' social behaviours. Furthermore, they indicate that smartphone digital phenotyping could be employed to provide personalised healthcare and maintain the QoL of wider communities.

6.1 Introduction

Long-term diseases have aroused increasing concern with the growing pursuit of quality of life (QoL). As Parkinson's disease (PD) is a common long-term disease and the number of patients is more than ten million, it is essential and valuable to study and observe the behaviour of PD patients. PD is an incurable long-term neurodegenerative disease with gradual loss of motor and non-motor functions. The motor symptoms of PD are apparent and easy to identify; they mainly include shaking, rigidity, slow movement and tremors. As for the non-motor symptoms, it is challenging to observe them directly; they involve mood changes, cognitive decline, pain, sleep disturbance, apathy and autonomic dysfunction [174]. All these symptoms significantly influence patients' social lives, which then causes social withdrawal. Practically speaking, social withdrawal indicates reduced social interactions. Interviews with PD patients also showed that their social connectedness was disrupted due to PD [203]. PD patients may lose and alter their social identities due to the disease's progression [206]. At the same time, they could also have social anxiety, a sense of embarrassment and less social confidence [206]. Because PD progression is idiosyncratic and people's social habits differ, PD patients have unique paths of social withdrawal [107]. Therefore, it is more reasonable to observe the social withdrawal of PD patients individually than across a population. However, to the best of our knowledge, research into social withdrawal in PD is still at an early stage. Researchers still rely on participants' experience and conceptual understanding to investigate this phenomenon.

Because it is a long-term disease, recognising the progression stages is essential for the treatment and management of PD. It can also help to maintain patients' wellbeing. Typically, clinical assessments are applied every six months to quantify patients' conditions [76]. However, PD symptoms fluctuate hourly and daily. The assessment could miss significant and detailed changes because it aims to cover a period of six months. Moreover, these assessments are possibly subjective, unreliable and biased [80], since they usually only rely on the experience and memory of patients or doctors. Due to these weaknesses, PD measurement needs a more continuous and objective approach. It would also be better if this approach were unobtrusive so that influence on patients' behaviour could be minimised.

Digital phenotyping is an emerging method for understanding human behaviour. It refers to using personal digital devices to quantify individual-level phenotype, moment by moment [164]. As a popular digital device, smartphones are used as a novel research tool here. It has been confirmed to be feasible and informative to use smartphones in this way [18], and the method promises to expand knowledge of PD patients' lives [19]. Almost everyone uses a smartphone regularly, and current smartphones have considerable abilities to conduct various tasks. Using smartphones to observe participants' behaviour is practical, unobtrusive and convenient without additional requests and costs. They play an essential role in people's social lives. People can use them to implement multiple communication methods, including messages, calls and social media. In addition, smartphones can infer face-to-face conversations, since they are embedded with sensors such as microphones that can detect the surrounding environment. After a simple initial configuration, smartphones can unobtrusively record data 24/7 without interruption. In this way, they can capture continuous rather than snapshot data. Therefore, not moment slices but the full picture of a patient's behaviours during a past period can be provided to clinicians. As every participant has their own smartphone, different PD patients' behaviour can be monitored individually, which is desirable for PD's idiosyncratic progression. In general, the smartphone is promising for monitoring overall social interactions individually. It can provide an objective understanding of social withdrawal.

Considering PD's long-term nature, a year-long longitudinal study was planned to observe social withdrawal among PD patients using smartphones. A designated application was installed on participants' smartphones to collect all possible social behaviour data and related factors. To explore the relationship between novel smartphone measures and existing clinical/psychological assessments, other standardised scales were also involved in the study. These assessments included clinical and psychological scales measuring QoL, PD progression and factors causing social withdrawal such as cognition, apathy and stigma. They were conducted every two months. In addition, a paper diary was provided to each participant to record their QoL and social ratings weekly.

With participants joining in and dropping out, eight participants with minor PD finished the year-long observation. In all, 23 features were built up across more than ten million raw smartphone data points. These features cover various aspects of social life, including calls, messages, social media usages, face-to-face conversations and locations. They were then correlated with all the assessments we conducted, including scales and diaries. The results show that our approach is able to reflect the social behaviour of participants. At least one smartphone feature for each participant was found to have a significant correlation with standard scales. The majority of the results (81%) implies that participants tend to have reduced social activities when situations are worse. Different strong correlations found for each participant also confirm that people's PD progression and social habits are distinctive. Furthermore, the whole procedure could also be applied to monitor other long-term health issues, enabling extensive understanding and the provision of targeted care.

6.2 Background

PD is an incurable long-term neurological disease with unknown causes. The symptoms of PD can be categorised as motor or non-motor. Motor symptoms are movement-related and include tremors, slow movement, rigidity, impaired posture and balance. Non-motor symptoms include neuropsychiatric, autonomic, gastrointestinal and sensory symptoms and sleep disorders [34]. All PD symptoms' progressions are fluctuated rather than linear [71]. Therefore, it is difficult to manage and treat PD caused by complex factors.

Symptoms of PD cause social withdrawal in a variety of ways. For motor symptoms like shaking, rigidity, bradykinesia and tremors, it is evident that they will impact PD patients' movement ability, which will make social engagement more difficult. The patients have to reduce the frequency of joining groups and taking part in social activities. Non-motor symptoms, including apathy, depression, stigma, social anxiety and cognitive impairment, can also cause reduced social interactions. It can be seen that apathy can reduce PD patients' motivation and passion for joining social events. Other non-motor symptoms, such as depression [213], stigma [139] and social anxiety [21], also have similar effects and ruin the experience of social interaction. In the case of cognitive impairment, the social functions of PD patients will be affected [108]. It will be hard for PD patients with cognitive impairment to follow conversations and understand others; they will then feel frustrated and maybe neglected by others [146]. Consequently, they will gradually become less interested in joining social activities because they cannot get pleasure from social interactions. In general, we can conclude that, on the one hand, PD can cause social withdrawal; on the other, social withdrawal is a potential indicator of symptom deterioration.

Social withdrawal signifies deviations of social behaviour from normal conditions, and it is evaluated by objective lack of contact with social network members [157]. It is different from loneliness, a subjective feeling that closeness of contacts does not satisfy people's minimum requirement. In our work, smartphones monitor participants' objective social behaviour unobtrusively and do not measure their subjective feelings. In other words, we focus on social withdrawal, which is quantifiable in terms of reduced social interactions. Many factors can cause social withdrawal, including diseases and sociodemographic factors. We cannot completely control all these factors when conducting a population-level experiment, so it is nearly impossible to apply a unified standard to estimate social withdrawal extent. Social withdrawal cannot be defined by a certain number of social activities being decreased; it can only be approached and studied at an individual level.

For a non-curable disease like PD, the focus of the treatment is to improve the QoL of the patients. According to the QoL questionnaire, the frequency and severity of non-motor symptoms have the most impact on QoL [141]. However, the importance of non-motor symptoms is often neglected [25]. Here, social factors are an essential part of QoL [246], and social functioning is found to be a particular interference when comparing the QoL of PD patients with the general population [194]. Therefore, social withdrawal, which is hugely impacted by non-motor symptoms, could indicate QoL.

6.3 Related Work

Smartphone Social Sensing Smartphones have already been widely used in researching human behaviour for decades. Using smartphones to collect human behaviour data is 'moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices', and this method is called smartphone-based digital phenotyping [164]. Smartphone social sensing has been applied in various fields, including general human social behaviour [239], personality effects [35], mental health (including depression, anxiety and stress) [119] and diseases like schizophrenia [28]). There are various types of social behaviour that can be captured or inferred via smartphones. Calls and messages are usually the first consideration. Contacts, length and incoming or outgoing status can be derived from calls and messages [195]. Social media usage is inferred from application usage [74]. Smartphones can also infer other interactions outside of themselves. Bluetooth can then be used to monitor people's proximity, since smartphones can continuously scan and search surrounding signals. As each person carries their own smartphone nowadays, a scan entry could indicate a person nearby [250]. Audio captured by microphones can be detected to infer nearby conversations [239]. All this information can be transformed into features and processed based on the research targets. For instance, when studying personality, correlations between collected smartphone data and the big five personality traits are built [237]. As a novel research tool, smartphones can observe participants' social behaviour continuously and unobtrusively, which is beneficial in social research.

Participants have no extra burden during data collection, and both contextual and behavioural data can be collected, so changes and deviations can be monitored from a more comprehensive perspective [11, 202].

Digital Phenotyping in PD This method of smartphone-based digital phenotyping is also applied in PD patients' behaviour studies. However, these studies mainly focus on the motor perception abilities of smartphones. Motor impairment in PD is classified by smartphone applications evaluating voice, posture, gait, finger tapping and response time [8]. Additionally, the accelerometer embedded in smartphones can be used to assess tremor intensity [117]. The strong correlation between the inferred results and clinical scale confirms that collecting data with smartphones is feasible for evaluating PD's severity. Although smartphones still play a significant role in these studies, some specific and intrusive tasks are requested to be done by patients. There is other research that takes advantage of the continuous monitoring ability of smartphones. For example, mobility features were generated from longitudinal passive smartphone data [94], and then the daily fluctuated motor symptoms of PD patients, including pain gait, freezing and fatigue, can be evaluated. PD patients' medication adherence was estimated by monitoring gait abnormality [260]. Finger tapping and memory tests were implemented on smartphones to detect longitudinal disease phenotypes [179]. Although smartphone-based digital phenotyping was applied in these studies, motor symptoms were their primary considerations, and these monitoring methods are usually intrusive. Other health indicators for PD patients, such as QoL and social impact, are still underestimated. To the best of our knowledge, PD patients' social withdrawal has never been studied using unobtrusive smartphone-based digital phenotyping.

6.4 A general model of social interaction

Social interaction is the basis of social lives, and it consists of a series of social behaviours with mutual communication. Critical factors in social interactions include a social target and relationships benefiting each other. Reduced social interaction is the key variable to be measured in social withdrawal. Alterations in the number of social interactions with time can indicate social changes. As social interactions rely on media to convey information, they have to take place over certain channels such as face-to-face communication or smartphone calls. By monitoring social interactions that happen in these channels, the comprehensive condition of social interactions can be established. For social contacts, psychological studies often focus on particular types of relationships. The well-known social brain hypothesis claimed that there is a limited capacity of contact that a human brain can process [56]. It divides the whole social network into four groups, and the number of relations starts at 5 and increases by a multiple of 3, but the intimacy level decreases. Specifically, there are support cliques (4–5), the sympathy group (12–15), the affinity group (around 50) and the active social network (around 150) [95]. Although why social relationships form these hierarchies is still unknown, this structure is consistent with common sense and has a significant influence on social relationship studies [220]. Overall, communication channels and types of contacts are two key components constructing social interactions, so the overall social behaviour will be measured by these two perspectives.

For smartphone social measurement, communication channels are differentiated by whether social interactions happening in that channel can be captured directly on the phone. For non-smartphone-mediated ones, it implies direct interactions, which is face-to-face conversations. The smartphone provides various kinds of communications for smartphone-mediated communications, including calls, messages and social media usage. However, based on the channel's information and emotion delivery approach, we believe there are existing ranks of these communication channels. The face-to-face level is the highest among all channels. Also, the effort for communicating in each channel is consistent with this rank. Face-to-face requires all participants involved to have immediate reactions; multiple brain functions have to operate to process both visual and verbal signals simultaneously and give appropriate responses. Although immediate reactions are still necessary for calls, only verbal cues need to be processed and reacted to. As for messages, instant replies are not required, so they take less effort for participants' brains to handle. This also applies to social media, which is more casual and does not require reciprocal responses. Combined with PD, which impacts the brain's social capabilities, more complex social interactions could indicate that the disease has slighter influences [168].

For social contacts, the social brain hypothesis has provided the concept of social groups. As the number of contacts in each group increases, the intimacy level decreases. These groups are termed support clique, sympathy group, affinity group and active social network. Although the hypothesis does not give a pragmatic definition of these terms, the groups' core difference is the familiarity of contacts, so we practically comprehend them as family, close friends, friends and acquaintances. Usually, family and close friends are people with whom one can share personal affairs. Closer relationships are also revealed by the acceptance of each other. People will feel less awkward when problems happen if they are more intimate. This also applies to PD patients. When their symptoms become severe, they could become anxious about social interactions. The first type of contact they withdraw from would be the most unfamiliar people, the strangers, followed by acquaintances and then friends. Therefore, social interactions with strangers are the strongest signals that people are socially active.

Based on the discussion above, two pyramids of social interaction importance can be constructed. Communication channels and types of contacts are incorporated to estimate the significance of a single social interaction in terms of PD. Details are shown in Figure 6.1. Their relative importance in terms of social withdrawal was represented by the size of the section in each pyramid. Face-to-face communication has the largest size in the pyramid of communication channels because it is a more noticeable indicator of participants' social functionality than other channels. Accordingly, calls ranked second, messages third and social media last. Likewise, social interactions with strangers are the most significant indicator that participants are socially active, so they have the largest size in the types-of-contact pyramid. The lines connecting the sections of the two pyramids convey the overall significance of social activities, as indicated by the number of dashes and the line's colour depth. The number of dashes and the line's colour depth can be compared independently. When two lines have the same colour, more dashes are less significant. Similarly, when two lines have the same dash levels, the lighter coloured line is less significant. Thus, the darkest no-dash line indicates that face-to-face conversation with strangers is the most significant indicator that people are not socially withdrawn. In contrast, social media with family members is the least significant, as seen by the lightest dashed line.

6.5 A year-long longitudinal experiment

As discussed in the previous chapters, the practical goal we plan to achieve is to use smartphones to measure social interactions among PD patients. Their social withdrawal could be subtle when observed over a short period, so a year-long observational study was initiated.

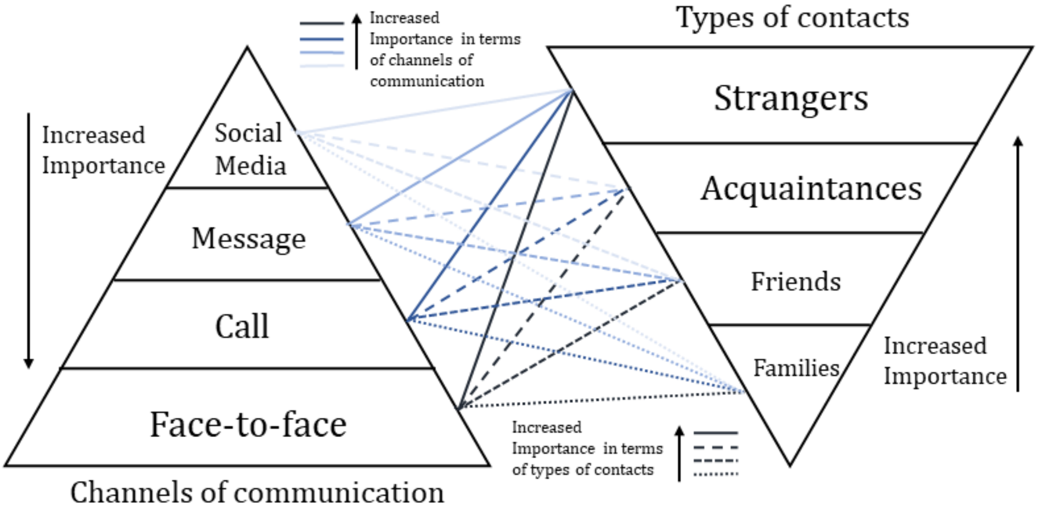


Figure 6.1: Two pyramids of social interaction importance.

A smartphone application called AWARE [67] was selected as a monitoring tool to be installed on participants' smartphones. It can capture various social-behaviour-related sensor data and social interactions that happen on smartphones 24/7. After an initial setup, it can also run in the background without interfering with participants, which meets our requirement of unobtrusive observation. Possible social-related data sources, such as calls, messages, phone usage, Bluetooth, GPS, keyboard and microphone, were monitored. The reasons for collecting these data are related to a systematic review of passive smartphone social sensing [257]. Types of data collected, their purposes and structures are listed in Table 6.1. All smartphone data collection was ethically processed. Identifiable entries like contact IDs and Bluetooth/Wi-Fi addresses/IDs were irreversibly encrypted. For sensitive information such as keyboard typing, only the length of characters typed was counted. The microphone was only used to detect whether there was a conversation at a given moment. This was achieved by implementing the algorithm developed in [239], and no raw audio was recorded. All these procedures were managed locally on smartphones and transmitted to a secure server we had physical control over.

Although we applied smartphones as a novel monitoring tool, other matured measurements were still necessary to validate the collected data. Following the paradigm found in the systematic review [257], a series of clinical and psychological scales were

Table 6.1: Data source, purposes and structures of collected sensor data [259].

Data source	Purpose	Structures
Calls	Call events	Timestamp, contact ID, length, status
Messages	Messages events	Timestamp, contact ID, status
Application usage	Time spent on social media	Open timestamp, app name, app package name
Notifications	Estimations of social media messages numbers	Timestamp, target application name
Bluetooth	Estimations of face-to-face encounters	Timestamp, Bluetooth address, Bluetooth ID
Wi-Fi	Estimations of locations	Timestamp, Wi-Fi address, Wi-Fi ID
GPS	Locations	Timestamp, longitude, latitude
Keyboard	Estimations of social media messages length	Timestamp, app name, app package name, length
Microphone	Detection of surrounding sound	Timestamp, is conversation

also included in the longitudinal study. The widely used Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [76] and Parkinson's Disease Questionnaire (PDQ-39) [102] were utilised to measure the progression of PD. As discussed in the background, several PD symptoms, including depression, stigma, apathy, empathy and cognitive impairment, are closely related to social withdrawal caused by PD. Accordingly, we picked the scale of each of them to obtain the state of these symptoms in each visit. Furthermore, we modified a social withdrawal scale for motor neurone disease (MND) to measure social withdrawal [185]. As another long-term neurodegenerative disease, MND has similar symptoms and impact on the QoL of participants, so the scale's validity is transferable to PD. We changed all descriptions of symptoms in the scale to Parkinson's to fit the aim of measuring social withdrawal in PD. All these scales were applied every two months to capture the state of these variables at that time. Per the requirement of MDS-UPDRS, all scales were conducted at participants' homes, that is to say, we visited participants' residences every two months. Participants' smartphones were also checked during home visits to maintain the quality of data. Details of questionnaires applied and their purposes are shown in Table 6.2. After selecting all scales, we found it could take too long to ask participants to finish them in one session, so only empathic concern and perspective-taking scales of IRI were included, as they are more closely related to social functions than the other scales.

Table 6.2: Scales applied every two months during the study [259].

Scale name	Purpose	Reference
MDS-UPDRS	Clinical Parkinson's progression	[76]
PDQ-39	QoL impact by Parkinson's	[102]
The Addenbrooke's cognitive examination revised (ACE-R)	Cognition	[147]
Stigma scale for chronic illness 8-item version (SSCI-8)	Stigma	[151]
Geriatric depression scale (GDS)	Depression	[252]
Interpersonal reactivity index (IRI)	Empathy	[46]
Apathy scale (AS)	Apathy	[211]
Social withdrawal scale (SWS) modified (see Appendix B)	Social withdrawal	[185]

A specially designed paper diary was also provided to each participant to track PD symptoms, progression and social interaction levels. Julio originally created this diary to trace day-to-day fluctuations of personal PD symptoms [234]. Although four other prototypes were tested, including Bluetooth, physical buttons, NFC and microcontrollers, the paper diary achieved the highest acceptance and compliance. In the original design, participants choose three symptoms that most impact their lives and record them daily. To obtain more granular ratings, a shorter version of the Parkinson's disease questionnaire, PDQ-8 [103], was also added to the diary every week. We also designed a weekly questionnaire to measure participants' social interaction levels. This questionnaire asks participants to rate their social interaction levels from 0 to 10, and the levels include different types of contact: family, friends, acquaintances and strangers. An overall rating is also requested. We followed the implications in the original design for all added items to maintain consistency. All ratings were accomplished by filling in the tiny dot of the corresponding number. An example of the answered weekly diary is shown in Appendix D. An extra page explaining types of contact was added to each diary. The instructions for reconfiguring the application are listed on the back side of that page for participants to refer to if it stops working. The whole page stretches out and can be folded as a bookmark for the diary (see Appendix C). To keep the diary simple and accessible, the time period covered by each diary was two months. The used diary was collected during each home visit, and a new one was provided to participants. In summary, our study has four-dimensional monitoring: 1) passive smartphone data collection 24/7; 2) daily symptom ratings; 3) weekly PDQ and social items; 4) two-monthly clinical and psychological scales. The full data collection plan and an example of the diary is shown in Figure 6.2.

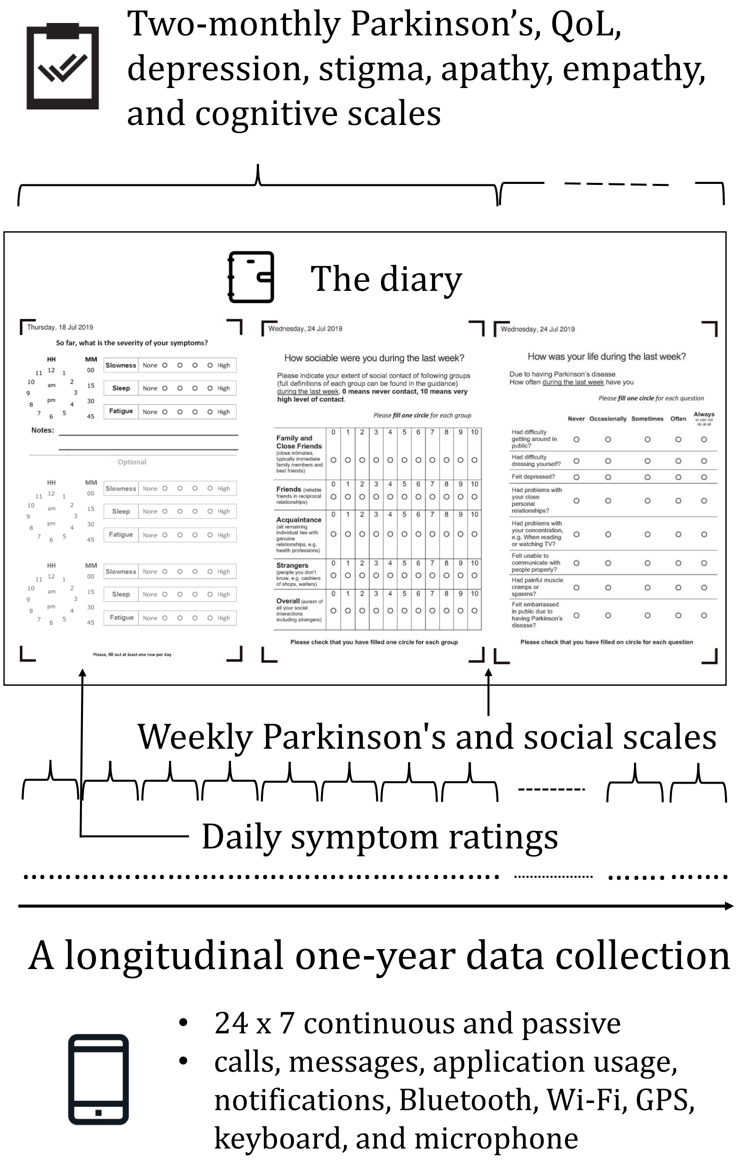


Figure 6.2: The plan for longitudinal one-year data collection and an example of the diary.

After the ethics approval, the participant recruitment campaign started. Our advertisements were distributed through the Parkinson's research community and the Parkinson's UK website. Potential candidates were also contacted via email and phone.

Since the experiment was administrated in the wild, not all participants needed to start simultaneously. Moreover, the only requirement for participants was that they were clinically diagnosed with PD and had the essential cognitive ability. All the recruitment took place on an enrol-and-go basis. Once the participant signed the agreement to join the study, the monitoring application was installed on their smartphone. Corresponding diaries were also provided to them. Most participants refused us, since it was a long commitment, so the recruitment took an extended period. Additionally, to eliminate the cost of accommodating novel environments, we installed the monitoring applications on participants' regular phones rather than providing them with secondary phones. Moreover, due to the restrictions of smartphones' operating system, iOS cannot provide essential social-related data: we only recruited participants who were Android smartphone users.

The whole experiment started in September 2019 and ended in March 2021. With participants joining in and dropping out, seven participants finished the whole year of observation. The COVID-19 pandemic severely impacted the experiment, not only theoretically but also physically. The planned home visits were interrupted after the official lockdown on 23 March. In addition, all scale-conducting sessions had to be moved online. Consequently, the full MDS-UPDRS could not be accomplished, since some items requesting touch and manoeuvres cannot be conducted via video calls. A similar situation happened with the cognition scale: it asks examinees to perform specific tasks in person, so the full scale cannot be achieved remotely. The smartphone stability check during home visits was also impacted. We were only able to check it remotely and send instructions to participants if they encountered any problems.

6.6 Results

The smartphone monitoring period and demographics of each finishing participant are shown in Table 6.3. The summary of raw data collected is shown in Table 6.4. As described in the previous section, we installed the application on participants' own smartphones. Although the compatibility of the phones was tested after the first installation, the monitoring software ran slowly and affected two participants' daily usage, so we provided them with new smartphones that could run it smoothly. Their data was transferred to these new phones, and necessary guidance was provided. They gradually adopted these new phones and treated them as primary phones. We also took a set of measures to maintain correct functioning of the application. A script was run to check data synchronisation every day. In addition to the page in the diary to help participants reconfigure the application, during the two-monthly home visit, we checked their smartphones manually to solve potential issues. These maintenance procedures continued online when home visits were suspended. Although every effort was made, 100% monitoring coverage is not guaranteed. The number of sensed days for P29 is significantly lower than for other participants, since the synced data is invalid.

Table 6.3: Monitored period and demographic of all participants.

Participant	Gender	Age	Start	End	Sensed days	Diary days
P23	F	65	Aug.30, 2019	Oct.10,2020	338	398
P24	M	73	Sept.26, 2019	Sept.2,2020	289	342
P25	M	76	Oct.17, 2019	Oct.16,2020	365	139
P26	M	75	Nov.18, 2019	Dec.16,2020	373	394
P28	F	63	Dec.5, 2019	Mar.10,2021	461	461
P29	M	66	Dec.2, 2019	Dec.23,2020	143	387
P31	M	78	Jan.15, 2020	June 7,2021	347	509
P32	M	64	Mar.12, 2020	June 23,2021	458	468

Table 6.4: Records collected for each smartphone sensor over all participants

Participant	P23	P24	P25	P26	P28	P29	P31	P32
Calls	668	1,003	1,576	499	486	207	1,470	1,552
Messages	1,733	855	1,275	412	3,673	243	1,197	1,183
Application usage	35,873	47,538	45,347	14,418	77,978	34,451	28,608	6549
Notifications	5,064	20,311	18,838	3,322	4,316	10,334	29,258	3,290
Bluetooth	312,289	450,725	202,766	696,554	942,500	1,875,492	882,698	298,318
Wi-Fi	2,027,030	1,702,975	2,020,057	5,679,420	42,993,870	279,948	517,985	4,289,646
GPS	848,477	116,076	309,538	630,824	475,068	216,268	361,402	118,661
Keyboard	254,196	41,246	32,666	5,147	58,389	44,624	67,012	32,365
Microphone	1,829,457	7,428,160	9,938,307	9,259,807	7,506,077	4,540,463	9,758,156	776,192

6.6.1 Compliance

For the questionnaires and diaries, a reasonable compliance rate was achieved. Overall, 100% questionnaire completion rate was achieved before the COVID-19 pandemic. Only two participants did not finish the last round of scales. The average compliance rate for the diary is 95%, the exception being one participant who did not return the diary for half a year. Specifically, the cognition scale contains the same tasks every time, and participants could be familiar with them if asked frequently, which affects the accuracy of the results; therefore, we reduced the frequency of ACE-R to every two visits. Even worse, after the lockdown, the cognition scale was abandoned because the majority of tasks could not be achieved online. Moreover, six items in the MDS-UPDRS that require touch and help from the examiner could not be performed after home visits were suspended. Therefore, these items were excluded when calculating the total score.

6.6.2 Pre-processing

Based on the raw smartphone data, we created a set of features related to social behaviours. All social interactions happen through specific communication channels. These channels are categorised into four main ones: calls, messages, social media and face-to-face. Calls, messages and social media happen on smartphones, so they can be captured directly. Bluetooth and microphone inferred face-to-face interactions. Three main characteristics describe a communication channel: frequency, duration and diversity. Practically, this means the communication channel's times, length and number of unique contacts.

For calls, these characteristics are clear. Times refers to the number of calls made, and both calls initiated and received were counted. Missed calls were excluded. Length is the total time of calls, and it counts in minutes to represent the time spent on calls. The unique contact is the number of non-repeating people called, and it indicates the size of call contact networks. As for messages, the length was replaced by the number of characters sent. We only considered sent messages because only they can indicate that the participants were involved in the interaction. This strategy also avoids one-way messages such as advertisements or announcements. The Android system does not record the number of typed characters for each message, so we innovated a strategy using the time the message was sent as a cut-off time to calculate the number of characters typed. Detailedly, the timestamp of the keyboard and the message sent are two different data sources. People usually finish typing their messages from the keyboard and send the message. So the number of characters they typed is likely to be the characters they sent in messages. Therefore, we use the timestamp of the message sent as the cut-off time to check the number of characters they typed at that time. Then the number of characters in the sent message can be known. The usage of other message applications such as WhatsApp, email and Skype was also considered. Unlike the messages in the system, the destinations of these messages cannot be known, but we can still combine the keyboard and application usage time to infer whether participants are using them. That is to say, if they are using the keyboard while message applications are in the foreground, a message is being sent. We also checked the application usage of each participant to ensure that every possible message application was included.

Similarly, social media usage was estimated by the time social media applications such as Facebook and Twitter were in the foreground. It was characterised by the times these applications were opened and the length of time they were used. Characters typed into these applications were also captured from the keyboard. All notifications of non-system messages, including both other message applications and social media, are quantified in notification times as another reference for the usage of these applications. Furthermore, as described in the background, a scanned Bluetooth signal could represent a person in close proximity, so we used the unique Bluetooth signals to estimate the number of people participants may have face-to-face social interactions with. The voice detection plug-in also gathered surrounding voices' start and end times, so the times and length of face-to-face interactions were extracted from each conversation.

Location factors of social interactions were also considered. From the systematic review [257], the length of time participants stayed at home, the times they left home, their travel distance, and the number of unique places they visited are important social factors. However, before establishing these features, raw GPS points must be transformed to semantic places. Therefore, a stop-point detection algorithm was adopted to identify whether the participants stayed at a place for a specific time [123]. Guided by previous research [99], we set our thresholds as 15 minutes and 100 metres. This means that, if the distance between GPS sequences is within 100 meters for 15 minutes, we believe participants started to stay in a single place. Since we visited participants' homes to conduct scales, the home coordinates were acknowledged for each participant. The home was then treated as a particular place, and the distance between each place visited and the home was calculated. Thus, we can finally compute the period participants stayed at home, the times they left home and the total travel distance between places and homes. The full features we constructed are shown in Table 6.5.

Table 6.5: Features created from social behaviour factors and their description.

Social Behaviour Factors	Features	Description
Calls	call times	Times of active calls
	call length	Total length of calls
	call unique	Number of unique contacts of calls
Messages	message times	Times of sent system messages
	message length	Total number of characters of sent system messages
	message unique	Number of unique contacts of sent system messages
	social message times	Times of sent non-system messages
	social message length	Total number of characters of sent non-system messages
	social message open length	Total time of non-system messages applications in foreground
	social message open times	Times of non-system messages applications opened
	social message notification times	Total times of non-system messages notifications
Social Media Usage	social media times	Times of active social media interactions
	social media length	Total number of characters typed in social media applications
	social media open length	Total time of social media applications in foreground
	social media open times	Times of social media applications opened
	social media notification times	Total times of social media applications notifications
Face-to-face	unique Bluetooth	unique Bluetooth signals (infer people in close)
	conversation times	Total number of face-to-face conversations
	conversation length	Total length of face-to-face conversations
Locations	home length	Total time spend at home
	out times	Total number of times left home
	out distance	Total distance travelled between visited places and home
	visited places	Total number of non-repetitive places visited

6.6.3 Correlation with scales

We conducted a correlation analysis of all participants and two-month scales. Since our visits were planned to be every two months, we asked participants to recall their situation changes in the last two months when answering questionnaires. However, the interval of these visits was not exactly two months every time. To keep the same length of time for each set of scales for comparison, we practically included eight whole weeks, which is 56 days' data. This also enabled weekly diaries to be involved in the next section. To maintain the integrity of 56 days, we tolerated a 10% loss of data, which means that, if the sensed time period was less than 52 days, the whole period was abandoned. This affects 8.11% of all the data, and the average number of sets of data included from an individual is 4.25. The correlation analysis is conducted between smartphone features listed in Table 6.5 and scales listed in Table 6.2. For the smartphone features, they were accumulated every 56 days. For example, for the call times, the number of phone calls of the 56 days before scales were conducted was correlated with the scales. At least three sets of data were included in the correlation analysis for each participant. To maintain consistency, all features of less than 56 days were scaled up to 56 days.

As discussed above, PD progression is idiosyncratic, and each participant has their preference. We expect that there will not be strongly correlated universal features across all participants. Thus, we conducted the interclass correlation analysis from the participants' perspective. As a higher score on these scales always means situations are worse, we expected that smartphone features would have negative correlations with these scales (except length of stay at home, which would be positive). The r ($-1 \leq r \leq 1$) indicates the strength and direction of the correlation. The *P*-value of each correlation pair was also calculated to examine the significance. The correlation results of every participant are shown in Tables 6.6 to 6.13. Only the *P*-value of a coefficient r less than or equal to 0.1 is reported.

As can be seen from the tables, most (81%) correlation results meet our expectations. Two features for each participant were found to have an anticipated correlation. This strongly indicates that there are connections between social behaviour and measured health and psychological conditions. This finding applies to all factors observed, including PD progression, QoL, empathy, apathy and stigma. In more detail, when participants' PD conditions, life quality, empathy, apathy or stigma levels are worse,

Table 6.6: Significant correlation between features and all scales of P23.

Scales	Features	r	*P*-value
SWS	call times	-0.935	*P*=.02
	call length	-0.821	*P*=.09
	social media length	0.815	*P*=.09
	unique Bluetooth	-0.857	*P*=.06
	conversation times	-0.879	*P*=.049
	conversation length	-0.928	*P*=.02
PDQ-39	call unique	-0.933	*P*=.02
	social media length	0.985	*P*=.002
	conversation length	-0.954	*P*=.01
	out times	-0.85	*P*=.068
MDS-UPDRS	social message open length	0.844	*P*=.07
	social media open length	0.990	*P*=.001
	out times	-0.878	*P*=.05
AS	social message open length	0.846	*P*=.07
	unique Bluetooth	-0.912	*P*=.03
	conversation times	-0.832	*P*=.08
SSCI-8	social message open times	0.906	*P*=.03
	social media notification times	-0.832	*P*=.01

Table 6.7: Significant correlation between features and all scales of P24.

Scales	Features	r	*P*-value
SWS	home length	-0.936	*P*=.06
PDQ-39	social message times	-0.974	*P*=.027
MDS-UPDRS	call times	-0.900	*P*=.1
	social message notification times	-0.913	*P*=.087
	out distance	0.976	*P*=.02
GDS	social message notification times	-0.918	*P*=.08
IRI-Sub	call times	-0.946	*P*=.05
	social message notification times	-0.957	*P*=.04
	out distance	0.942	*P*=.058
AS	unique Bluetooth	-0.94	*P*=.06
SSCI-8	social message notification times	-0.907	*P*=.09

there is a possibility that they have fewer social activities.

SWS directly measures the social withdrawal of participants. Theoretically, it has the closest relationship with number of social activities among the included scales. A

Table 6.8: Significant correlation between features and all scales of P25.

Scales	Features	r	*P*-value
SWS	unique Bluetooth	-0.923	*P*=.076
	conversation times	-0.972	*P*=.028
	conversation length	-0.959	*P*=.04
	home length	0.943	*P*=.057
GDS	unique Bluetooth	-0.943	*P*=.066
AS	message length	-0.914	*P*=.085
	message unique	-0.941	*P*=.059
	unique Bluetooth	-0.970	*P*=.03
	conversation times	-0.955	*P*=.045
	conversation length	-0.959	*P*=.04
	home length	0.989	*P*=.01
	out times	-0.958	*P*=.04
	visited places	-0.990	*P*=.01

Table 6.9: Significant correlation between features and all scales of P26.

Scales	Features	r	*P*-value
SWS	social message times	-0.903	*P*=.036
	conversation length	-0.922	*P*=.026
MDS-UPDRS	message length	-0.892	*P*=.04
	social message times	-0.965	*P*=.01
	social message length	-0.908	*P*=.03
	social message notification times	0.912	*P*=.03
	conversation length	-0.912	*P*=.03
IRI-Sub	call times	-0.947	*P*=.01
	call unique	-0.911	*P*=.03
	message length	-0.842	*P*=.07
	social message length	-0.846	*P*=.07
	social message notification times	0.905	*P*=.03
SSCI-8	call times	-0.855	*P*=.06
	message length	-0.832	*P*=.08
	social message times	-0.884	*P*=.047
	social message length	-0.846	*P*=.07
	social message notification times	0.936	*P*=.019

higher SWS number indicates participants are more socially withdrawn, which means they have reduced social activities. The correlation results affirmed this assumption. At least one feature was found to be strongly correlated in SWS for each participant. This is across all communication channels, including calls, messages, conversations

Table 6.10: Significant correlation between features and all scales of P28.

Scales	Features	r	*P*-value
SWS	call times	-0.905	*P*=.03
	call unique	-0.879	*P*=.05
	call length	-0.943	*P*=.016
	message length	-0.808	*P*=.098
	social message times	-0.895	*P*=.04
	social message open length	-0.806	*P*=.1
	social message open times	-0.823	*P*=.087
	social media times	-0.964	*P*=.01
	social media length	-0.930	*P*=.02
	social media notification times	-0.886	*P*=.045
	social media open length	-0.848	*P*=.07
	social media open times	-0.929	*P*=.02
	home length	-0.884	*P*=.05
PDQ-39	call length	0.919	*P*=.027
	social media times	0.916	*P*=.03
	social media open times	0.954	*P*=.01
	social media open length	0.942	*P*=.017
	unique Bluetooth	-0.940	*P*=.017
MDS-UPDRS	call times	0.991	*P*=.001
	call unique	0.980	*P*=.003
	social message open times	0.966	*P*=.008
	social media length	0.997	*P*<.001
	home length	0.977	*P*=.004
IRI-Sub	social message notification times	-0.823	*P*=.087
SSCI-8	message times	-0.879	*P*=.05
	social message length	-0.844	*P*=.07

and social media usage. Significant correlations were also found in location factors, including home length, out times, visited places and out distance. The majority of them meet our expectations, except the home length of P24 and P28. At least one channel of communication from smartphone data is associated with SWS for each participant except P24. This could support the idea that collected smartphone social behaviour data can reflect participants' social withdrawal. Participants are likely to have decreased social activities when SWS total scores are higher.

Although other scales did not involve every participant, this provides evidence that

Table 6.11: Significant correlation between features and all scales of P29.

Scales	Features	r	*P*-value level
SWS	call unique	0.887	*P*=.04
	social message times	-0.922	*P*=.03
	social message length	-0.963	*P*=.009
	social message notification times	-0.938	*P*=.018
	social message open times	-0.975	*P*=.005
MDS-UPDRS	call unique	0.835	*P*=.079
	message unique	-0.816	*P*=.03
	social message times	-0.893	*P*=.009
	social message length	-0.977	*P*=.009
	social message notification times	-0.869	*P*=.009
	social message open times	-0.98	*P*=.009
GDS	social message open length	-0.937	*P*=.02
	unique Bluetooth	-0.993	*P*<.001
	home length	0.930	*P*=.02
IRI-sub	call unique	0.926	*P*=.02
	message unique	-0.973	*P*=.05
	social message times	-0.897	*P*=.039
	social message length	-0.905	*P*=.035
	social message open times	-0.930	*P*=.02
	social media notification times	0.859	*P*=.06
AS	unique Bluetooth	-0.813	*P*=.09
	out distance	-0.852	*P*=.067

Table 6.12: Significant correlation between features and all scales of P31.

Scales	Features	r	*P*-value
SWS	call length	-0.886	*P*=.046
MDS-UPDRS	out times	-0.817	*P*=.09

Table 6.13: Significant correlation between features and all scales of P32.

Scales	Features	r	*P*-value level
SWS	call length	-0.903	*P*=.097
PDQ-39	call unique	-0.941	*P*=.059
	unique Bluetooth	-0.980	*P*=.02
	visited places	-0.995	*P*=.005
GDS	call length	-0.977	*P*=.02
	message unique	-0.913	*P*=.087

these factors may have associations with social activity levels captured by a smartphone. We assumed that these correlations would be similar to SWS, which are negative because the deterioration of these factors could reduce participants' social abilities. The majority of results are consistent with our expectations. There exist strong negative correlations between smartphone features and the total score on scales. These results suggest that, with increased levels of PD progression, apathy, stigma and decreased level of empathy, particular participants tend to have fewer social activities. However, 19% of the results do not meet our expectations. These results involve the measurement of communication channels and location factors. They could indicate other novel phenomena. For example, there is a strong positive correlation between PDQ-39 and social media length in P23. This indicates that, when PD gets worse, P23 tends to type more characters in social media apps. We will discuss the possible explanation for these results in the discussion section.

6.6.4 Correlation with diaries

Based on sets of 56 days of data, we also conducted correlation analysis with diaries. All ratings included PDQ-8, and social ratings were added to eight weeks. The same sets of smartphone data were correlated with eight weeks' sum of PDQ-8 and overall social ratings. We employed the same correlations to these sets of measurement and smartphone data. Since a higher score of PDQ-8 indicates participants have worse QoL, we expected it would have negative correlations with smartphone features. Overall social ratings and contact type ratings were expected to have the same trend as smartphone features, as these ratings directly measure the social activity level of participants. Similarly, home length is the reversed one: positive correlations with PDQ-8 and negative correlations with social ratings were expected. As with the correlation with scales, only P-value ess than or equal to 0.1 are reported.

Intimacy levels were differentiated for each communication channel to achieve the pyramid of communication in Section 6.4. We adopted the results from previous studies on people' perception of intimacy from smartphone context data [85, 84]. These have several conclusions: 1) unique contact is always the strongest indicator of intimacy for known contacts' communications, and the most contacted number has the highest intimacy; 2) for face-to-face interactions, fewer people around equals higher levels of intimacy; 3) lower intimacy tends to have shorter sessions when using applications.

Table 6.14: Significant correlation between features and diaries of P23.

Sections	Features	r	*P*-value
PDQ-8	social message times	-0.986	*P*=.015
	social message length	-0.944	*P*=.057
	social media times	-0.902	*P*=.098
	social media notification times	-0.929	*P*=.07
Overall social ratings	social message open length	-0.954	*P*=.01
	unique Bluetooth	0.957	*P*=.01
	home length	-0.892	*P*=.04
Friends social ratings	group 2 conversation times	0.866	*P*=.058
	group 2 conversation length	0.898	*P*=.038
Strangers social ratings	group 4 call times	0.874	*P*=.05

Table 6.15: Significant correlation between features and diaries of P24.

Sections	Features	r	*P*-value
PDQ-8	social message open length	-0.914	*P*=.086
Overall social ratings	social message open length	-0.935	*P*=.065
	unique Bluetooth	0.983	*P*=.018
	home length	-0.931	*P*=.069
	out times	0.913	*P*=.087
	visited places	0.974	*P*=.026
Friends social ratings	group 2 conversation times	0.942	*P*=.058
Acquaintance social ratings	group 3 conversation times	0.952	*P*=.048
	group 3 conversation length	0.970	*P*=.03
Strangers social ratings	group 4 conversation length	0.948	*P*=.05

The first conclusion was applied to calls and messages, since their unique IDs can be known. The number of Bluetooth signals could infer the number of people around. Therefore, for each face-to-face conversation, we recorded the number of unique Bluetooth signals as its intimacy reference. A lower number of unique Bluetooth signals could represent a higher level of this face-to-face conversation. In order to adopt the third conclusion, we summed the session of each social media application usage from the foreground application time stamps. By scanning the foreground application log, the length of sessions can be known by the start and exit time difference.

After an intimacy reference was generated for each communication channel, we applied K-means to differentiate them into four groups. K-means is a popular clustering

Table 6.16: Significant correlation between features and diaries of P26.

Sections	Features	r	P-value
PDQ-8	social media open length	0.939	P=.018
Overall social ratings	social message open length	0.746	P=.089
	conversation length	0.879	P=.02
	home length	-0.875	P=.02
	out times	0.903	P=.01
	out distance	0.752	P=.08
	visited places	0.896	P=.016
Acquaintance social ratings	group 3 conversation times	0.792	P=.06
	group 3 conversation length	0.82	P=.046
Strangers social ratings	group 4 call length	0.889	P=.018
	group 4 message times	0.895	P=.016
	group 4 message length	0.895	P=.016
	group 4 conversation length	0.881	P=.02

Table 6.17: Significant correlation between features and diaries of P28.

Sections	Features	r	P-value
PDQ-8	call length	0.842	P=.035
	unique Bluetooth	-0.921	P=.009
Overall social ratings	message times	0.828	P=.04
	social message length	0.758	P=.08
	visited places	0.807	P=.05
Friends social ratings	group 2 message times	0.864	P=.026
	group 2 message length	0.865	P=.026
Acquaintance social ratings	group 3 call times	0.861	P=.028
	group 3 social media times	0.852	P=.03
	group 3 social media length	0.861	P=.028
Strangers social ratings	group 4 call times	0.865	P=.026
	group 4 call length	0.969	P=.001
	group 4 conversation times	0.878	P=.02
	group 4 conversation length	0.878	P=.02
	group 4 social media times	0.878	P=.02
	group 4 social media length	0.878	P=.02

mechanism that could partition observations into designated clusters. Each observation is assigned to the cluster with the nearest mean. K-means fitted our aim because we planned to categorise all communication in single channels into four groups. The intimacy reference we created has numerical features. For example, families could be

Table 6.18: Significant correlation between features and diaries of P31.

Sections	Features	r	*P*-value
PDQ-8	call unique	0.991	*P*=.08
Overall social ratings	social message length	-0.992	*P*=.002
	out distance	0.997	*P*=.007
Acquaintance social ratings	group 3 call length	0.999	*P*=.001
	group 3 conversation times	0.998	*P*=.039
Strangers social ratings	group 4 call length	0.999	*P*=.002
	group 4 conversation times	0.988	*P*=.099
	group 4 conversation length	0.988	*P*=.099

Table 6.19: Significant correlation between features and diaries of P32.

Sections	Features	r	*P*-value
Overall social ratings	out times	0.671	*P*=.099
Acquaintance social ratings	group 3 message times	0.680	*P*=.09
	group 3 message length	0.688	*P*=.087
Strangers social ratings	group 4 call length	0.757	*P*=.049

contacted many times in a period of time, but strangers are only contacted once. As a result, the means of families and strangers are distinctive and can be easily clustered by K-means. Eventually, eight features, including calls, messages, social media and face-to-face conversations, were created for each group. Groups 1–4 correspond to family, friends, acquaintances and strangers, respectively. A prefix was added in front of these features to distinguish them from non-intimacy-related features.

P25 did not return enough diaries, and P29 did not have enough qualified data, so these two participants were excluded from the analysis. The results show that significant correlations have been found in all six eligible participants, as shown in Tables 6.14 to 6.19. Significant correlations between smartphone features and PDQ-8 were found in five participants but not P32. The majority of results reflect our assumptions, but three correlations are positive. This suggests participants could increase communication in specific channels if QoL deteriorates. The call length correlation with PDQ-8 of P28 is the same situation with PDQ-39, so P28 is likely to spend more time on phone calls when QoL worsens. In addition, as we expected, significant correlations between overall social ratings and smartphone features were discovered in every participant. Like the SWS results, this supports the idea that smartphone data could reflect participants' social behaviours. Although not all four contact types are covered in each participant, at least two contact types were found to be significantly correlated with the corresponding group smartphone features. All four communication channel features exist in the significantly correlated contact types ratings. This demonstrates the ability of our approach to practically differentiate the intimacy of social behaviours.

6.7 Discussion

From the results, correlations of all participants across all scales were taken. For each participant, at least two smartphone features were found to have a significant correlation with these scales. Therefore, smartphone social data has the potential to reflect the situation of all these scales. Digital phenotyping provides another perspective to understand all related social factors of PD, including stigma, empathy and apathy. This can be a reference for understanding social withdrawal and the course of PD. The impact and effectiveness of interventions for PD can also be potentially reflected by digital phenotyping. As we hypothesised, there is no single feature correlated through different participants. Moreover, none of the participants found significant correlations in all conducted scales. This indicates that every participant has their own social habits and PD progression path. As illustrated in the background, social behaviours and even disease progressions are affected by many factors, and it is impossible to control all these factors to have a unified study. Social withdrawal in PD is a complex phenomenon that is better understood individually.

As can be seen from the correlation results, significant correlations between SWS and smartphone features exist in every participant. SWS has the closest relationship with social behaviours, as it directly measures the extent of social withdrawal. It is reasonable that SWS has the widest significant correlations with smartphones compared with other standardised scales. Moreover, different factors of social withdrawal play different roles in correlations with smartphone data. Scales other than SWS have significant correlations with smartphone data in all participants, which confirms the discussion in the background that social withdrawal is a complex phenomenon that is influenced by various factors such as depression, apathy, empathy and PD progression. For all participants, the smartphone data were found to have a significant correlation with PDQ-39 or MDS-UPDRS, which has the highest number of correlated participants among all scales except SWS. This could indicate PD progression, and QoL could influence social withdrawal and probably have more influence than other factors. As for the other scales, at least four participants' smartphone data were found to have significant correlations with them. These results also exemplified that social withdrawal is an individualised phenomenon and that various factors influence every participant's social withdrawal. However, the smartphone features significantly correlated with GDS were not always location factors. This is not consistent with [188], which used smartphones to observe the depression of participants and found that location factors were related to depressive symptom severity. It could also be caused by the scales we used. GDS is usually used for elderly participants, and [188] applied Patient Health Questionnaire-9 (PHQ-9), a more general questionnaire for depression. Moreover, we focused on individual level correlation, but [188]'s results are based on all 40 participants.

All created features existed in significant correlations with scales, so all these features were potentially applicable to the study of social withdrawal and related factors. The smartphone feature also exists at different times for specific scales. If a particular smartphone feature has been found to be significantly correlated with multiple scales of a single participant, that could indicate that this feature has important implications for that participant. For example, message length negatively correlated with MDS-UPDRS, IRI-Sub and SSCI-8 in P26, so messages are essential communication channels for P26, which can be an important indicator of P26's social situation. More attention can be given to P26's messaging behaviour to study their social withdrawal. Most of the correlation results meet our expectations that participants tend to have reduced social activities when situations are worse. However, few correlations indicate that participants are likely to increase social interactions when scale scores are higher. For example, P28 was found to have increased call length when PDQ-39 was higher. Similarly, P28's MDS-UPDRS had a positive correlation with unique calls and call times, which indicated this participant tends to make phone calls to more contacts when PD worsens. These results could be attributed to participants' social habits and personal preferences. One possible explanation of these positive correlations is that participants tend to communicate with others about their worse situations. They need comfort, relief or help from their support network, and when they have difficult feelings, they express them more with others, which causes an increase in certain communication channels. The strong correlation also reveals preferences for communication channels. For example, P31 and P32 only have a significant call length correlation with SWS. Nevertheless, four out of five of P29's correlations with SWS are message-related. This could imply that when social withdrawal happened, P31 and P32's calls and P29's messages could be more evident than other communication methods.

Although there were not as many of them as there were significant correlations in scales, correlations in diaries also demonstrated the potential of the in-depth understanding of social behaviours. For the five participants with a significant correlation of diaries, at least two contact types per participant were found to correlate significantly with the corresponding group's smartphone features. This suggests that smartphones could achieve more detailed levels of social interaction monitoring, considering both communication channels and contact types. Correlations with groups also potentially reveal participants' communication preferences for different contact types. For example, significant correlations between messages times/length and friends and between social media times/length and acquaintances were found in P28. This indicates that P28 may prefer to communicate with friends through messages but with acquaintances through social media. Therefore, reduced message-based communication potentially indicates that P28 is socially withdrawn from friends. Overall, not only the correlations themselves but also other findings, such as individual communication preference, reactions towards worse situations and in-depth observation of social activities, can be retrieved from the digital phenotyping. It is a novel approach and provides objective and original knowledge of people's social behaviour.

6.8 Limitations and Future Work

To the best of our knowledge, this is the first study applying these scales at such high-frequency intervals and applying digital phenotyping to study social withdrawal in PD. Although the study has the advantage of novelty among studies of its kind, there are still limitations that can be further considered.

COVID-19 is a significant factor that impacted the experiment, and measures have been taken to alleviate these impacts. During each questionnaire answering session, we asked all participants only to consider the impact of PD. For items within the scales, participants were asked to answer from the perspective of PD only. However, even with these measures, the social restrictions caused by COVID-19 cannot be ignored. It would be intriguing to see the results under typical conditions where people can have social interactions freely. These results could be compared with our findings to investigate how COVID-19 impacts PD patients longitudinally.

We are a proof-of-concept study. As we mentioned in the background, social withdrawal can be influenced by various factors. Measured variables are not guaranteed to cover factors that influence social withdrawal. Social behaviour changes that happened to participants can be attributed to various factors. Other dedicated experiment designs could be implemented to discover how PD impacts social withdrawal clinically. Our approach demonstrates its potential for an in-depth understanding of social activities considering both communication channels and contact types; however, not all participants find strong correlations with all contact types. One possible reason for this is that the technique for categorising contact types from smartphone data is still at an early stage. More smartphone features for differentiating types of contact could be studied to make multiple reason decisions rather than just relying on the frequency of contact; the contact types of each participant could then be more reasonably categorised to give a comprehensive understanding of each social activity.

Since it is a long-term disease, the progression of PD can take years, and a year-long study may be just the tip of the iceberg. Minimal clinically important difference (MCID) is the smallest change of scales that a patient identifies as meaningful clinical change [97]. By comparing the summary index of PDQ-39 at the beginning and end of the experiment, only four participants appear to have MCIDs. Two of them even indicated their QoL was better at the end. This implies that a year-long observation of PD patients may be still unlikely to capture significant clinical progression. Therefore, a more extended experiment should be considered to reveal the prolonged clinically significant progression of PD and the social withdrawal of participants. For granularity of observation, two months could still be too long to provide participants' latest situation. More detailed social behaviour change could provide more granular data to investigate social withdrawal. With increased granularity, personalised social behaviour models can be established from smartphone data.

6.9 Conclusion

This paper presents the results of a longitudinal study of social withdrawal in PD. It applies smartphone digital phenotyping technology and related standardised scales to observe social activities among PD patients. The strong and significant correlations between smartphone data and scales provide evidence that smartphones could reflect the social withdrawal and related factors of PD participants. Although possible positive associations between PD progression and smartphone-monitored social withdrawal were discovered, the phenomenon is complex and needs additional exploration. This kind of monitoring could provide another perspective for understanding social withdrawal in PD. It has the potential to be a reference for PD progression and QoL individually, which could be used in wider communities to promote personalised health services.

Chapter 7

Tracking Social Behaviour

The previous chapter demonstrates relationships between psychological/clinical scales and smartphone data. The majority of them indicate that social interactions are reduced when situations are worse. However, these scales were conducted every two months, which is still probably not enough for fluctuated progression. Apart from correlation, another aim of our study is to construct a personalised model for each individual in a more granular manner. So, their social interaction level and quality of life can be known from smartphone data rather than self-report scales. This chapter is the expansion of the previous chapter and shares the same background and experiment. All collected smartphone data are analysed towards weekly ratings in this chapter. From the overall conceptual model, we consider both the level of social ratings and social level changes, so that both numerical and directional predictions can be achieved. The techniques for differentiating types of contact are also attempted. Different subset separations of the whole data are examined to test the performance of the generated models.

The content of this chapter is adapted from *Heng Zhang, Bijan Parsia, Ellen Poliakoff, and Simon Harper. 'Tracking Social Behaviour with Smartphones in People with Parkinson's: A Longitudinal Study'*. It's currently under view.

Author's contributions

Heng Zhang designed and conducted the research, collected the data, analysed and synthesised the findings, and wrote the paper. Ellen Poliakoff gave suggestions on study design and helped to conduct the study. Bijan Parsia and Simon Harper provided constant input throughout the study, including advice, feedback and critical revisions to the final manuscript.

Abstract

Parkinson's disease (PD) is a chronic neurological disease that has both motor and non-motor symptoms that negatively influence patients' quality of life (QoL). Social life is a substantial part of QoL, and reduced social interactions are a result of both PD symptoms and impairment in QoL. This is known as social withdrawal, and it can be a sign of general health in persons with PD. Smartphones are suitable for monitoring social behaviour under the notion of digital phenotyping, as people spend a large amount of time socialising on them. We implement year-long longitudinal research using smartphones to study social withdrawal in PD. In addition to an entire year of 24/7 continuous monitoring on smartphones, weekly diaries of social ratings and QoL are provided by participants as self-report ground truth. Twenty-three features are extracted from more than 10 million raw data entries collected from eight participants. These features are then used to build models to reflect participants' self-report ratings. We consider the interactions that happen on smartphones, such as calls, messages and social media, and use smartphone sensors to infer face-to-face interactions. By applying multiple linear regression and Naive Bayes, our model achieves at least 0.6 R-squared in numerical prediction and a 0.6 F1 score in the direct projection of all participants. It is significantly better than just assigning mean and random guesses. The results suggest that our approach provides a more granular method for tracking social behaviour in people with PD via smartphones, and it could benefit wider communities where the social impact on patients needs attention.

7.1 Introduction

Parkinson's disease (PD) is a prevalent incurable long-term neurodegenerative disease, and its symptoms can cause reduced social interactions in PD patients [178]. This phenomenon is termed social withdrawal. The goal of treating patients with a non-curable disease like PD is to improve their quality of life (QoL). As social interaction is a significant part of QoL [194], social withdrawal could be an overall indicator of PD progression. Tracking social withdrawal could provide another perspective to understanding PD and provide PD patients with better health services. However, to the best of our knowledge, researching social withdrawal in PD is still at an early stage.

Alternatively, recognising PD progression is a prerequisite for treating and managing

PD and maintaining patients' QoL. These evaluations still rely on traditional clinical and psychological scales, so the results could be influenced by the experience and memory of PD patients or clinicians [80]. Moreover, the patients are in an aware situation when conducting these scales, so Hawthorne effects could be introduced [197]. Overall, traditional measures are likely to be subjective, unreliable and biased. Accordingly, novel PD measurement requires continuous, objective and unobtrusive monitoring, which also applies to tracking social withdrawal in PD patients.

With the popularity of digital devices, digital phenotyping has become a novel method used to observe human behaviours. Objective behavioural information can be captured continuously by digital devices [164]. Almost everyone has a smartphone, so collecting information from a smartphone is convenient without interrupting participants, and it has been confirmed to be a reasonable and informative method [18]. It also has the potential to study PD patients' lives [19]. Smartphones have become an important communication tool, and they can sense the surrounding social context of the user. Therefore, the social withdrawal of PD patients may be determined by analysing the data from smartphones.

Above all, as a long-term disease, social withdrawal in PD patients should be recognised and analysed on a long-term basis. We implemented year-long longitudinal research to track PD patients' social withdrawal using smartphones. With people joining in and dropping out, eight participants with minor PD completed the whole year of study. Data were collected using an installed application on participants' smartphones. The application always ran in the background and did not ask participants to complete any extra tasks on their smartphones. All factors related to social behaviour, including messages, phone calls, social media usage, face-to-face conversations and locations, were collected continuously. We also provided a paper diary for each participant to record their QoL and social ratings every week as self-report ground truth.

Finally, more than 10 million raw smartphone data points were obtained. Twenty-three features covering different measures of social interactions and related factors were extracted from the raw data to describe participants' social behaviour. Channels of communication and types of contacts were considered to construct the social behaviour model. Based on the concept of numerical predictions and direction of changes, we applied multiple linear regression and Naive Bayes to establish social behaviour models for diary ratings. Data, and subsets of data, were split to test the performance of the models. For numerical prediction, our model achieved at least 0.6 R-squared, and a 0.6 F1 score was achieved in the direct projection of all participants. This result indicates that our model successfully reflects participants' social interaction and QoL levels from smartphone data on a weekly basis. These models were established explicitly for particular participants to reveal the individual differences between participants. Our method provides a personal and in-depth understanding of the social behaviour of PD patients, and it could help doctors and carers of patients with PD to perceive their QoL and disease progression in a more granular manner. This method could also be promoted to broader health communities where the social impact on patients needs attention.

To summarise, our contributions are threefold:

- **A feature extraction mechanism for retrieving social behaviour from raw smartphone data (Section 7.6.2)** presents a set of behavioural metrics that reflects levels of social interactions of each participant. With smartphone-based passive sensing, these metrics can be gathered unobtrusively over a long period of time.

- **A social behaviour model to track social withdrawal in PD patients (Section 7.4)** considers various social elements, including all social interactions mediated by smartphones, face-to-face conversations, locations and types of contacts.

- **An evaluation of both numerical and direction models (Section 7.6.4 and Section 7.6.5)** validates our approach against participants' ground truth. The results suggest our models successfully track participants' social behaviour, outperforming just assigning mean and random guesses.

7.2 A primer on social withdrawal in PD

PD causes social withdrawal PD is an incurable cause-unknown long-term neurological disease that affects over 10 million patients worldwide. The degenerative disorder in the central nervous system of PD patients causes the brain to decline. PD has two types of symptoms, motor and non-motor [221], both of which can cause reduced social interactions in PD patients [178]. With shaking, rigidity, bradykinesia and tremors, it is difficult for PD patients to move; therefore, they might decrease the frequency of going out and participating in social activities [216]. Non-motor symptoms can also cause reduced social interactions. Apathy [173], depression [213], stigma [139] and social anxiety [21] can reduce PD patients' motivation and passion for interacting socially. The social functions of PD patients are also impacted by cognitive impairment [108]. The status of people who have reduced social interaction is termed social withdrawal, which is evaluated by the lack of social contact with other people [157]. As all of these PD symptoms cause social withdrawal, it could be an overall indicator of PD.

Social withdrawal could be an overall indicator Compared to the general population, social functioning is an important interference in the QoL of PD patients [194], and it has been reported that PD patients lack social confidence [203]. As an incurable disease, all treatments aim to improve the QoL of PD patients, and sufficient monitoring is a prerequisite of providing appropriate treatments. However, clinical assessments are typically conducted every six months [76], which may be inadequate for a fluctuated disease [140]. More granular monitoring could provide a more sophisticated understanding of PD. Furthermore, the progression of PD is idiosyncratic [200], and every patient has a unique path of development [107]. This also applies to PD causing social withdrawal, and as other sociodemographic factors can also influence social behaviours, it is more reasonable to understand it on an individual basis. By measuring social withdrawal, both PD progression and QoL can be reflected.

7.3 Related work

Smartphone social sensing As a hub of personal communication, a significant amount of social interaction happens on smartphones. Therefore, miscellaneous social information can be gathered from smartphones for different research purposes. The approach of 'moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices' is termed smartphone-based digital phenotyping [164], and it has been applied in various psychological and clinical fields, including personality [182], depression [236], stress [115] and diseases like bipolar disorder [66]. Collecting smartphone data provides a novel viewpoint from which to understand these subjects. Social interactions mediated by smartphones, such as calls, messages and social application usage, can be directly captured from the system. As people always carry smartphones, face-to-face interactions could also be inferred from embedded sensors, for example, a Bluetooth signal could represent a person nearby.

With continuous scanning, a smartphone equipped with Bluetooth could infer that people are in proximity [250]. Microphones can also detect surrounding sound to infer conversations [239]. More importantly, as a unique research instrument, smartphones can monitor participants' social behaviour continually and unobtrusively [193]. There is no additional strain on participants during data collection, and both environmental and behavioural data are captured [53]. Therefore, smartphone social sensing could provide a comprehensive perspective to observe changes and deviations in behaviour [32].

Digital phenotyping in PD Digital phenotyping has also been used to study PD patients' behaviour, but the majority of studies focus on motor symptoms. Typically, participants are asked to perform tasks on a smartphone application, involving voice, finger tapping, gait, balance and reaction time, and the results are used to classify PD motor severity [255]. Furthermore, the accelerometer embedded in smartphones can sense tremor intensity, so participants with different levels of rigidity and bradykinesia can be identified [117]. However, these studies need participants to complete extra work, which is intrusive and may be burdensome. Other studies take advantage of smartphones' continuous monitoring ability to conduct longitudinal experiments. Mobility features generate longitudinal passive smartphone data that are used to monitor fluctuated motor symptoms, including pain, gait, freezing and fatigue [94]. Tasks like finger tapping and memory tests are also longitudinal so that PD progression phenotypes can be learnt [179]. Nonetheless, motor symptoms are still the primary consideration of these studies. Other impacts of PD, including non-motor symptoms, QoL and social withdrawal, still need to be explored. To the best of our knowledge, we are the first to use unobtrusive smartphone-based digital phenotyping to study PD patients' social withdrawal.

7.4 Social interaction model

Practically, social withdrawal is defined as reduced social interaction [233]; therefore, social interaction is the variable to be measured. The social interaction model is a structure that describes and reflects the strength of social interaction. As a mutual behaviour, social interactions are conveyed to a specific target through certain media, so communication channels and types of social contacts are two aspects that quantify social interactions.

Communication channels that can be analysed through smartphones are divided into smartphone-mediated and non-smartphone-mediated categories. In smartphone-mediated communication channels, social interactions are directly implied by phone calls, messages and social media usage. For non-smartphone-mediated communication channels, direct interactions (i.e. face-to-face conversations) can be inferred from Bluetooth and the microphone. Furthermore, the social affordances level differs in distinct channels [73], so we believe that all of these channels can be ranked according to the brain function involved. Face-to-face is the highest because participants have to process both visual and verbal signals and give immediate reactions. Phone calls rank second because only an immediate verbal response is required. Messages rank third because signals do not have to be processed instantly. Social media is more casual than the previous three channels and does not need reciprocal responses, so it ranks fourth. PD influences brain function, therefore, it will also impact social interactions. If a PD patient can still execute more complicated social interactions that involve higher brain functions, then PD has a lesser impact [168].

In terms of types of social contacts, the social brain hypothesis [56] claims that due to the limit of the human brain, an entire social network is divided into four groups: support clique (4–5 people), sympathy group (12–15 people), affinity group (around 50 people) and the active social network (around 150 people) [95]. The hypothesis does not give a pragmatic definition of these groups, but intimacy is the main difference between them. Therefore, we practically define these four groups as families, friends, acquaintances and strangers. PD patients may become awkward and anxious about social interactions when symptoms worsen. As they are more trusting of families and friends, they will only maintain communication with them and withdraw first from people who are unfamiliar to them (i.e. strangers). Thus, if PD patients still socially interact with strangers, it is a strong signal that they are still socially active; however, if they do not interact with family and friends, they may have severe social withdrawal.

From the discussion above, a model for describing the importance of social interaction is established, which is represented in the two pyramids in Figure 7.1. In each pyramid, the section sizes indicate the rank of importance in terms of social withdrawal. For communication channels, face-to-face interactions are a more evident signal than other channels that participants are socially functional; therefore, it is the largest section in the pyramid. Similarly, social activities with strangers are a more significant cue that participants are socially active. Lines between the sections of the two pyramids represent overall social activity significance, which is reflected by the number of dashes and the line's depth of colour. A lighter colour line with more dashes means that the social activity is less significant, considering the communication channel and types of contacts. Therefore, it can be understood from the figure that face-to-face communication with strangers is an indicator of more positive social interaction, shown by the darkest non-dashed line, and vice versa for social media with families as being the least important, shown by the lightest dashed line.

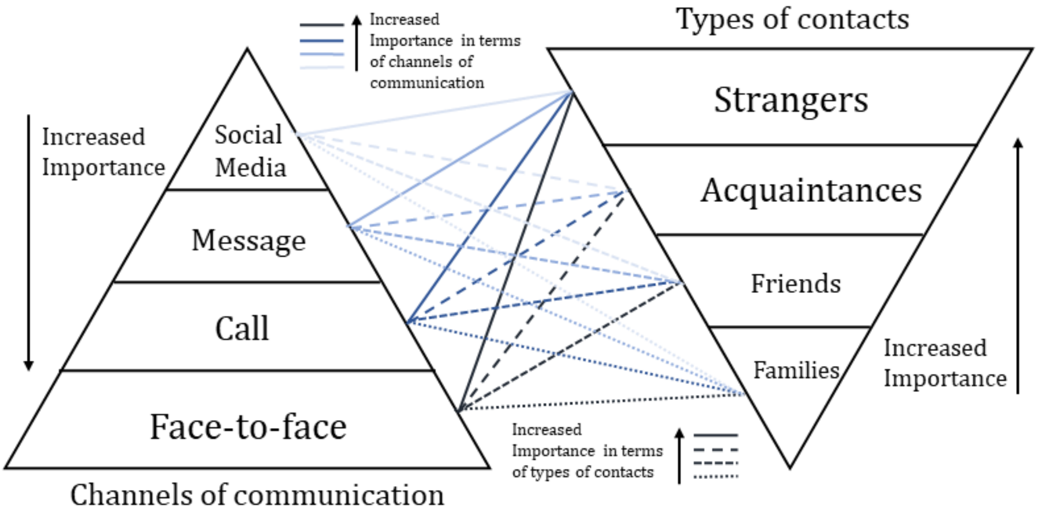


Figure 7.1: Two pyramids showing the importance of social interaction [258]

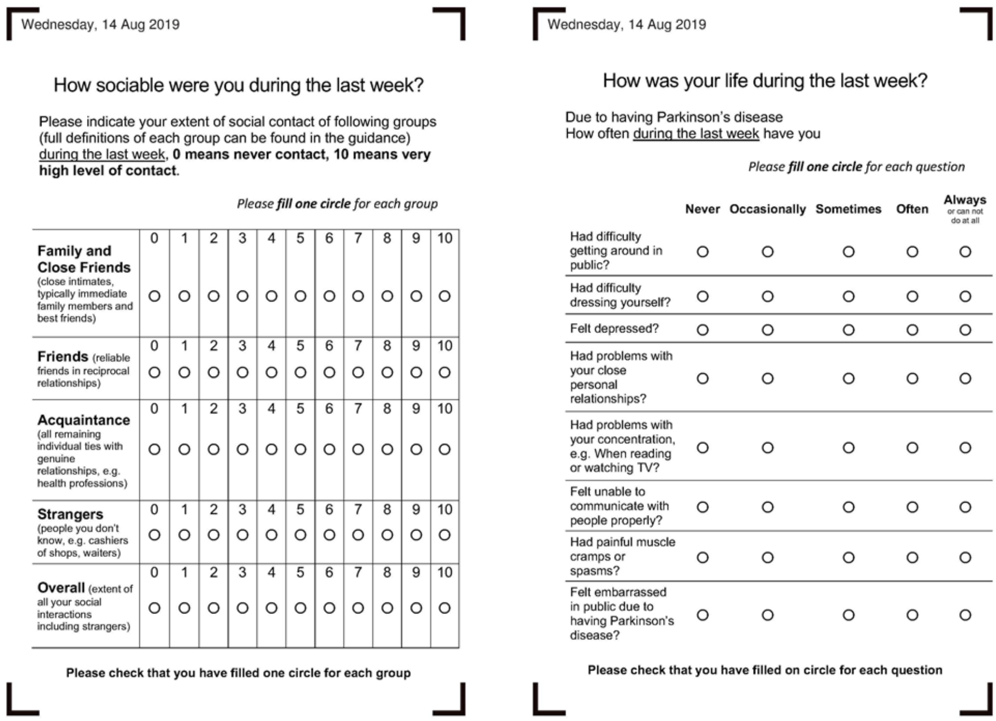
7.5 Methods

To achieve the practical goal of monitoring social withdrawal among PD patients, a year-long observational study using smartphones was initiated. A data collection platform named AWARE [67] was selected to be installed on participants' smartphones. This application ran unobtrusively 24/7 to collect all social-related information as mentioned in [257]. The details of the collected raw data, the purpose and data structures are given in Table 7.1. All collected smartphone data were considered ethically, and identifiable entries were irreversibly encrypted. Only the length of the typed characters on the keyboard was recorded, and the microphone only detected if there was a conversation [239]. All data were transmitted to a secure server that we could physically control. Apart from smartphone data collection, a set of clinical and psychological

Table 7.1: Data source, purposes and structures of collected sensor data [259]

Data source	Purpose	Structures
Calls	Call events	Timestamp, contact ID, length, status
Messages	Messages events	Timestamp, contact ID, status
Application usage	Time spent on social media	Open timestamp, app name, app package name
Notifications	Estimations of social media messages numbers	Timestamp, target application name
Bluetooth	Estimations of face-to-face encounters	Timestamp, Bluetooth address, Bluetooth ID
Wi-Fi	Estimations of locations	Timestamp, Wi-Fi address, Wi-Fi ID
GPS	Locations	Timestamp, longitude, latitude
Keyboard	Estimations of social media messages length	Timestamp, app name, app package name, length
Microphone	Detection of surrounding sound	Timestamp, is conversation

scales were conducted every two months as references. The details of these scales are described elsewhere [258]. In addition to these scales, we provided a specially designed paper diary for each participant to track their QoL and social interaction levels weekly. The diary included a shorter version of the PD questionnaire, PDQ-8 [103], to measure QoL, and a weekly questionnaire asked the participants to rate their social interaction level from 0 to 10. Different social contacts were rated respectively, including family, friends, acquaintances and strangers. Following the design implication from Julio [234], all ratings were made by placing a tiny dot on the corresponding number, which made the diary accessible for PD patients. Figure 7.2 shows a detailed example of the diary.



Wednesday, 14 Aug 2019

How sociable were you during the last week?

Please indicate your extent of social contact of following groups (full definitions of each group can be found in the guidance) during the last week, **0 means never contact, 10 means very high level of contact**.

*Please **fill one circle** for each group*

	0	1	2	3	4	5	6	7	8	9	10
Family and Close Friends (close intimates, typically immediate family members and best friends)	O	O	O	O	O	O	O	O	O	O	O
Friends (reliable friends in reciprocal relationships)	O	O	O	O	O	O	O	O	O	O	O
Acquaintance (all remaining individual ties with genuine relationships, e.g. health professions)	O	O	O	O	O	O	O	O	O	O	O
Strangers (people you don't know, e.g. cashiers of shops, waiters)	O	O	O	O	O	O	O	O	O	O	O
Overall (extent of all your social interactions including strangers)	O	O	O	O	O	O	O	O	O	O	O

Please check that you have filled one circle for each group

Wednesday, 14 Aug 2019

How was your life during the last week?

Due to having Parkinson's disease
How often during the last week have you

*Please **fill one circle** for each question*

	Never	Occasionally	Sometimes	Often	Always or can not do at all
Had difficulty getting around in public?	O	O	O	O	O
Had difficulty dressing yourself?	O	O	O	O	O
Felt depressed?	O	O	O	O	O
Had problems with your close personal relationships?	O	O	O	O	O
Had problems with your concentration, e.g. When reading or watching TV?	O	O	O	O	O
Felt unable to communicate with people properly?	O	O	O	O	O
Had painful muscle cramps or spasms?	O	O	O	O	O
Felt embarrassed in public due to having Parkinson's disease?	O	O	O	O	O

Please check that you have filled on circle for each question

Figure 7.2: An example of diaries given to participants, social interaction levels on the left, PDQ-8 on the right

7.6 Results

7.6.1 Participant recruitment

The participant recruitment campaign started in September 2019, and the experiment ended in March 2021. We recruited participants who were clinically diagnosed with PD and who did not suffer from symptoms that severely impacted their usage of smartphones. As each participant was observed individually, recruitment was on an enrol and go basis. Once participants signed the agreement, the AWARE application was installed on their smartphones, and the diary was provided to them. Unfortunately, the iOS operating system did not provide essential social-related data, so only PD patients with Android smartphones were recruited. With participants dropping out for various reasons, eight participants finished the whole-year monitoring.

Table 7.2 shows the smartphone monitoring duration and demographics of every participant. Sensed days in the table are summarised from continuous monitoring sensors, such as GPS, Bluetooth and Wi-Fi. The week both scales finished is counted as a diary week. An example of the answered weekly diary is shown in Appendix D. The summary of the captured raw data is shown in Table 7.3. We implemented a series of safeguards to ensure the application's continued functionality. Every day, a script ran automatically to confirm data synchronisation. Participants' smartphones were checked manually during home visits or online to resolve any problems. An instruction was added to the diary to help participants reconfigure the application if needed (see Appendix C). While every effort was made, complete monitoring coverage cannot be assured; for instance, because the synced data were faulty, P29 had a much smaller number of sensed days than the other participants. As for diaries, a reasonable compliance rate was achieved. The only problem was that participants sometimes forgot to complete whole entries or scale entries for some weeks. We also verified social rating answers by comparing contact ratings and overall ratings. If the overall rating increased but every sub-rating decreased, or vice versa, that week's social ratings were deemed invalid. Except for one participant who did not return the diary for six months, the average diary compliance rate was 95%.

Table 7.2: Monitored period and demographic of all participants

Participant	Gender	Age	Start	End	Sensed days	Diary weeks
P23	F	65	Aug.30, 2019	Oct.10,2020	338	56
P24	M	73	Sept.26, 2019	Sept.2,2020	289	48
P25	M	76	Oct.17, 2019	Oct.16,2020	365	19
P26	M	75	Nov.18, 2019	Dec.16,2020	373	56
P28	F	63	Dec.5, 2019	Mar.10,2021	461	65
P29	M	66	Dec.2, 2019	Dec.23,2020	143	55
P31	M	78	Jan.15, 2020	June 7,2021	347	72
P32	M	64	Mar.12, 2020	June 23,2021	458	66

7.6.2 Feature Extraction

We constructed a set of features relating to social behaviours based on the raw smartphone data. All monitored social relationships took place via specific communication channels (i.e. calls, messaging, social media and face-to-face). Calls, messaging and social media interaction took place on smartphones, allowing them to be recorded in real-time. Face-to-face interactions were inferred using Bluetooth and the microphone. Frequency, length and diversity were the three main characteristics that described a

Table 7.3: Records collected for each smartphone sensor over all participants [258]

Participant	P23	P24	P25	P26	P28	P29	P31	P32
Calls	668	1,003	1,576	499	486	207	1,470	1,552
Messages	1,733	855	1,275	412	3,673	243	1,197	1,183
Application usage	35,873	47,538	45,347	14,418	77,978	34,451	28,608	6549
Notifications	5,064	20,311	18,838	3,322	4,316	10,334	29,258	3,290
Bluetooth	312,289	450,725	202,766	696,554	942,500	1,875,492	882,698	298,318
Wi-Fi	2,027,030	1,702,975	2,020,057	5,679,420	42,993,870	279,948	517,985	4,289,646
GPS	848,477	116,076	309,538	630,824	475,068	216,268	361,402	118,661
Keyboard	254,196	41,246	32,666	5,147	58,389	44,624	67,012	32,365
Microphone	1,829,457	7,428,160	9,938,307	9,259,807	7,506,077	4,540,463	9,758,156	776,192

communication channel and referred to the time, length and unique type of social interaction that took place on a communication channel over time.

As an example, in terms of calls, frequency was the number of calls, length was the total time of calls, and the number of non-repeatable phone numbers called was referred to as unique contacts. The number of characters sent replaced the length of messages. As the Android system does not keep track of the number of characters typed in each message, we innovated by summarising the number of characters typed using the time the message was sent as a cut-off time. Only sent messages were considered because they could determine that participants were active during message conversation. One-way messages, such as advertisements and announcements, were also filtered by this method, and other messaging apps, such as WhatsApp, email and Skype, were taken into account. Due to the system's limitation, unique contacts in these messaging apps were not revealed; however, we could still infer if participants were utilising them by combining keyboard and application usage time. If participants used the keyboard when message apps ran in the foreground, they were sending messages. Similarly, foreground application records could reveal social media usage. The number of times social media applications, such as Facebook and Twitter, were launched and the length of time they were used were filtered from these records. The keyboard was also used to capture the number of characters input and calculate the active usage times of these applications. Moreover, the number of notifications from non-system messages and social media applications was retrieved as another usage measurement.

Unique Bluetooth signals were used to estimate the number of people participants might have contacted face-to-face. The start and end times of nearby voices were also recorded by the voice recognition plug-in. So, from each face-to-face conversation, the times and lengths were extracted. Additionally, location factors affecting social interactions were considered. According to the systematic review, the length of time spent inside and outside the home, distance travelled and the number of places visited, were four significant social elements [257]. However, raw GPS points had to be translated into semantic locations before these properties could be established. Thus, whether the participant stayed at a particular location for a specified amount of time was determined by a stop point recognition algorithm [123]. We chose a threshold of 15 minutes and 100 metres based on prior research [99]. That is, if the distance between GPS sequences was less than 100 metres for 15 minutes, participants remained in one location. Then, the distance between each visited place and home was calculated. Thus, four significant social elements were achieved. The complete set of constructed features is provided in Table 7.4.

7.6.3 Intimacy Construction

To create the communication pyramid in section 7.4, distinct levels of intimacy were assigned to each communication channel. In practice, we used smartphone context data to replicate the findings of earlier studies on people's perceptions of intimacy [85, 84]. The following conclusions were reached: 1) For communications with known contacts, the strongest indicator of intimacy is always the unique contact, and the number contacted most frequently has the most intimacy; 2) For face-to-face conversations, fewer individuals in proximity means higher intimacy; 3) When using applications, people with a lower level of intimacy have shorter sessions.

Because unique IDs were known, the first conclusion was applied to calls and texts. The quantity of Bluetooth signals could be used to estimate the number of individuals in proximity. As a result, we utilised the number of distinct Bluetooth signals for each face-to-face chat to indicate intimacy. Fewer unique Bluetooth signals could represent a higher level of intimacy for this face-to-face interaction. We aggregated each social media application usage session from the foreground application timestamps to reach the third conclusion. The start and exit time differences in the foreground application log could be used to determine the length of sessions.

After generating an intimacy reference for each communication channel, they were divided into four groups using K-means, which is a prominent clustering algorithm

Table 7.4: Features created from social behaviour factors and their description [258]

Social Behaviour Factors	Features	Description
Calls	call times	Times of active calls
	call length	Total length of calls
	call unique	Number of unique contacts of calls
Messages	message times	Times of sent system messages
	message length	Total number of characters of sent system messages
	message unique	Number of unique contacts of sent system messages
	social message times	Times of sent non-system messages
	social message length	Total number of characters of sent non-system messages
	social message open length	Total time of non-system messages applications in foreground
	social message open times	Times of non-system messages applications opened
	social message notification times	Total times of non-system messages notifications
Social Media Usage	social media times	Times of active social media interactions
	social media length	Total number of characters typed in social media applications
	social media open length	Total time of social media applications in foreground
	social media open times	Times of social media applications opened
	social media notification times	Total times of social media applications notifications
Face-to-face	unique Bluetooth	unique Bluetooth signals (infer people in close)
	conversation times	Total number of face-to-face conversations
	conversation length	Total length of face-to-face conversations
Locations	home length	Total time spend at home
	out times	Total number of times left home
	out distance	Total distance travelled between visited places and home
	visited places	Total number of non-repetitive places visited

that can divide data into distinct groups. The cluster with the closest mean was assigned to each observation. Because we proposed categorising every communication in single channels into four groups, K-means met our goal. Additionally, the intimacy reference had numerical characteristics; for example, participants may contact their families multiple times over a certain period, whereas strangers were only contacted once. As a result, families' and strangers' numerical averages were distinguishable and suitable for clustering by K-means. Eight features (i.e. length, times) were eventually developed for each group, including calls, messages, social media and face-to-face conversations.

7.6.4 Rebuild weekly ratings

Our method aimed to re-establish participants' levels of social interaction. As the overall social ratings in the diary were treated as ground truth, the practical target was to generate ratings from the collected objective smartphone data. As discussed in the background section, a particular channel was necessary for all social interactions; therefore, we categorised them as face-to-face, call, message and social media. Theoretically, the level of social interactions should be the combination of social activities in all four channels as shown in the formula below:

$$Total\,Social\,Interaction\,Level = Face-to-face+Calls+Messages+Social\,media$$

Nevertheless, the reality is much more complex than the theory. First, overall social ratings are still a type of subjective questionnaire, and it is nearly impossible that every channel of communication is equally treated when participants complete their ratings. Participants may have their own preferences when they rate their overall social interaction levels, and some channels may have more weight than others from a personal perspective. For example, participants may give more weight to face-to-face interactions because these consume more energy than messaging. Second, there are different measures, even for one channel of communication, and we do not know which one took a more significant role when participants quantified it. Typically, a communication channel could be measured by length, time or unique contact, but participants could give more attention to one of these measures. For example, participants may think their call activities increase due to the number of growing contacts, but the total length of call decreases. In summary, the overall ratings are still the combination of social activities in all four channels but with different weights, and one particular variable of each channel is selected. Considering that location is also a critical factor in social interactions, we also included it in the formula as shown below:

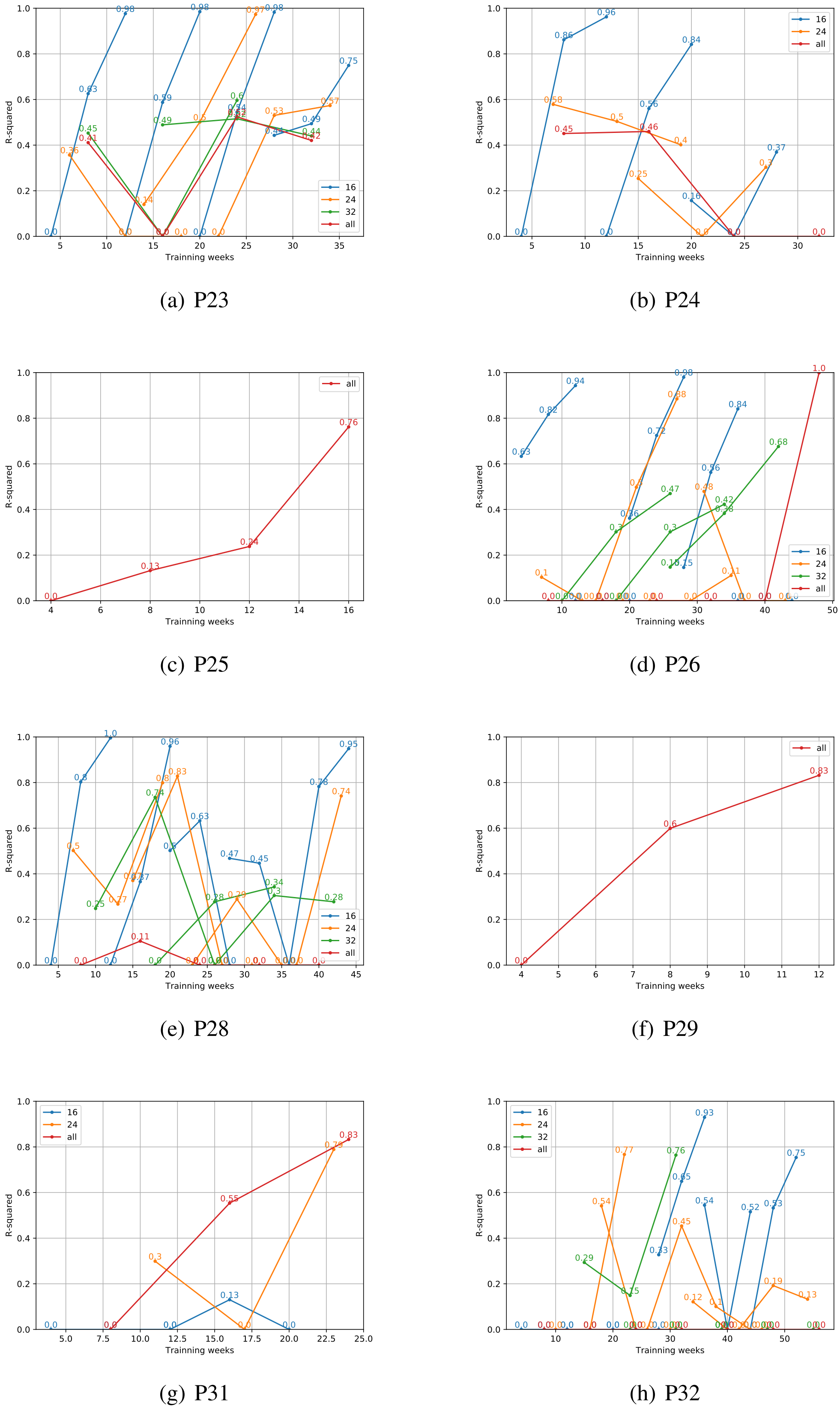
$$Total\,Social\,Interaction\,Level = \alpha_1 * Face-to-face\binom{Variables}{1} + \alpha_2 * Calls\binom{Variables}{1}$$
$$+\alpha_3 * Messages\binom{Variables}{1} + \alpha_4 * Social\,media\binom{Variables}{1} + \alpha_5 * Location\,media\binom{Variables}{1}$$

As we aimed to allocate proper weights to communication channels and obtain participants' social ratings, linear regression was a satisfactory fit for the target as it can represent multiple explanatory factors to a scalar by assigning distinct parameters. By minimising the difference between observed and predicted values, a link between the independent variables and the target variable will be discovered. Therefore, by feeding linear regression with smartphone data and social ratings, weights of communication channels can be computed. Then, the relationships between participants' weekly social ratings and smartphone data are established, which is in line with the formula above. The data input for training the linear regression was exclusively from the participant, so the relationship between smartphone features and social ratings established by linear regression will overfit the participants. This results in the generation of a customised model. As discussed in the background section, the progression of PD could impact the social lives of PD patients; therefore, we also applied this model to the PDQ-8 summary index using the same feature sets. Likewise, the ratings for different groups of contacts could be aggregated from features of that group in each communication channel. For example, the formula of the family ratings is shown below:

$$Family\,Social\,Ratings = \alpha_1 * Family\,face-to-face\binom{length, times}{1} +$$
$$\alpha_2 * Family\,calls\binom{length, times}{1} + \alpha_3 * Family\,messages\binom{length, times}{1} +$$
$$\alpha_4 * Family\,social\,media\binom{length, times}{1}$$

As the feature was divided by weeks, it was more practical to only count the number of sensed days more than five as a valid week. Therefore, the number of weeks included in this stage for each participant was P23:40, P24:34, P25:19, P26:50, P28:49, P29:17, P31:29 and P32:63. The whole dataset and different parts of 16, 24 and 32 weeks, representing four, six and eight months, were included for each performance test procedure. The different parts of the weeks were selected from rolling strategies. For example, a dataset of 40 weeks was split into weeks 0 to 16, weeks 8 to 24, weeks 16 to 32 and weeks 24 to 40. For participants whose included weeks were less than these weeks, we only considered the maximum possible parts. The training and testing set was also split from each test procedure by three percentages. For example, for 24 weeks, the number of training weeks was six, 12 and 18 weeks, so the testing weeks were 18, 12 and six weeks, correspondingly. As for the whole dataset, the training set started at eight weeks and increased by eight weeks each time until the end. R-squared values less than 0 were treated as 0 because it indicated that the generated model gave more errors than just assigning the mean. The highest R-squared value in this particular train/test split is reported in Figure 7.3 for each participant. Sixteen, 24 or 32 in the legend denotes 16, 24 or 32-week parts of the whole dataset. This indicates that all of the participants' data were included in the train/test procedure. The point on the figure indicates the highest R-squared value achieved by this number of training weeks.

As Figures 7.3 show, a reasonable R-squared is achieved across all participants. For the highest R-squared value of all participants, the minimum was 0.76, and the maximum was 1. This implies that our model outperforms the just assigning mean model with appropriate training weeks. Generally, the performance of the model increased with the number of training weeks, which is evident from the 16-week data part. For example, in P23's first 16 weeks, R-squared starts at 0 at four weeks, increases to 0.63 at eight weeks and reaches 0.98 at 10 weeks. The explanation for this could be that if more weeks are included in the training, the model can use more details of social behaviour patterns to make better predictions. However, there are also exceptions where R-square fluctuates with an increased number of training weeks. For example, the minimum R-squared appears in the last eight weeks of P24's final 16-week data part. In some circumstances, the highest R-squared is 0 at all times in all data parts, which is exhibited by P32's total data parts. Typically, the highest R-squared is achieved in the 16-week data part. The R-squared of the whole dataset as a train/test procedure are unusually low, except for P25 and P29 because their included weeks were limited. This could be caused by the number of training weeks, as the whole dataset usually has a higher test/train ratio compared with the 16-week data part. Furthermore, the changes in R-squared value driven by the increased number of training weeks is different. For example, in P24's first 16 weeks, R-squared grows from 0.86 to 0.96 because of four weeks incremental training weeks. However, in the second part of the 16 weeks,



(a) P23 (b) P24

(c) P25 (d) P26

(e) P28 (f) P29

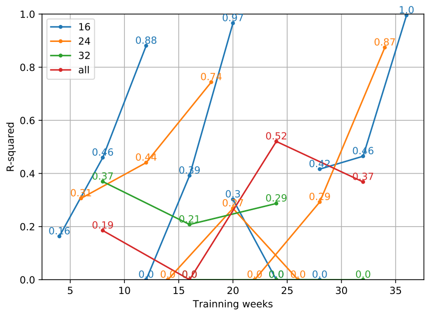
(g) P31 (h) P32

Figure 7.3: R-squared value of the multiple linear regression model applied to every participant's weekly social ratings

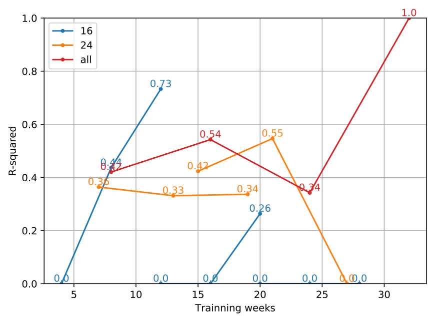
the four-week increment causes R-squared to rise from 0.56 to 0.84. As drawn from the discussion above, different participants have diverse situations, and every participant has distinctive R-squared values in all data parts. This could be explained by the uniqueness of the social patterns of each individual.

The same train/test procedure was conducted on every participant's weekly PDQ-8 summary index as shown in Figure 7.4. Similar to overall social ratings, reasonable R-squared is also achieved across all participants. The minimum R-squared was 0.6, and the maximum was 1 out of the highest R-squared values of all participants. The results also indicate that compared with the just assigning mean model, our model outperforms with appropriate training weeks. The model's effect on overall social ratings also applies to the model of the weekly PDQ-8 summary index. In general, performance increased with the number of training weeks; for example, across all of P28's 16 weeks. Nevertheless, R-square fluctuated with the increasing number of training weeks in some cases. In particular, the P31 model only exceeded the mean in parts of the 16 weeks. The effects of increasing the number of training weeks on R-squared was also inconsistent. Similar to P28's second part of the 16 weeks, R-squared remained 0.33 with the increment from eight to 12 training weeks. As PDQ-8 measures the QoL of participants, they also retain the individuality of their lives, so these results confirm the uniqueness of each individual and explain distinctive R-squared values.

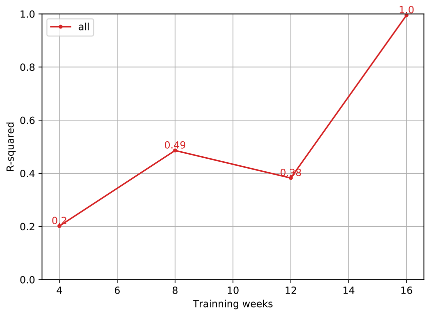
The model of different types of contacts applied the same data split procedure as the models of overall social ratings and PDQ-8. The median and range of of each type of contact of every participant is shown in Table 7.5. As illustrated in the table, our model outperforms the just assigning mean model at least once in every type of contact for every participant. This shows the potential of our model to reflect the social extent of different types of contacts. However, the number of R-squared values above 0 was much less than overall social ratings and PDQ-8, and the highest R-squared value was still less than those in the overall social ratings and PDQ-8 model. This is because the intimacy model was based on higher-level features to differentiate the types of contacts. It could induce more errors than relying on first-level features, such as the overall ratings and PDQ-8 models.



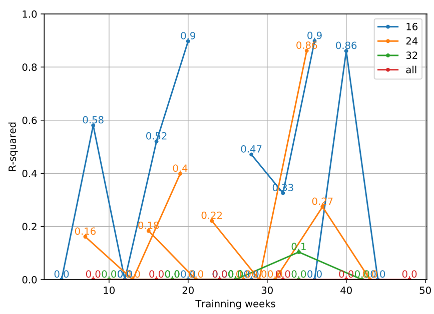
(a) P23



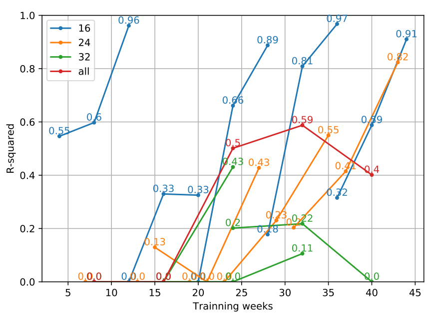
(b) P24



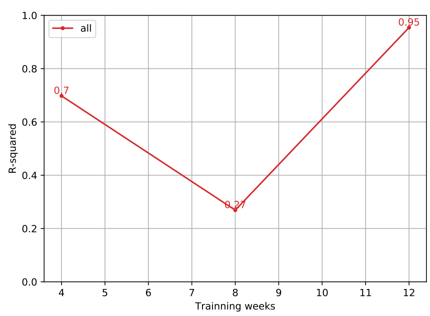
(c) P25



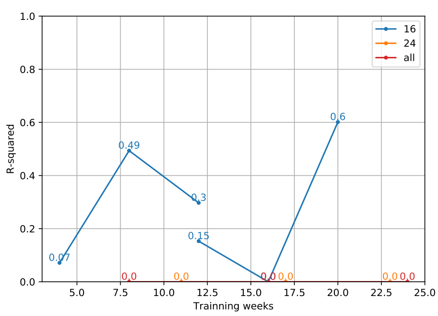
(d) P26



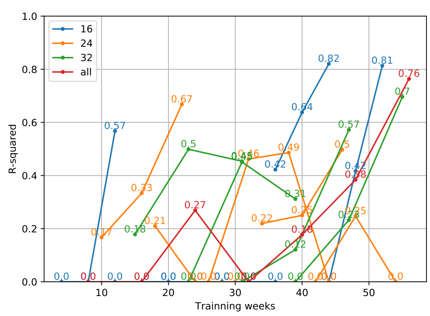
(e) P28



(f) P29



(g) P31



(h) P32

Figure 7.4: R-squared value of the linear regression model in every participant's PDQ-8 summary index

Table 7.5: Median and range of R-squared of every contact type of each participant achieved by the linear regression model

Participant	Contact types	R-squared Range	R-squared Median
P23	Family	(0.93, 0)	0.3
	Friends	(0.30, 0)	0.06
	Acquaintances	(0.52, 0)	0
	Strangers	(0.03, 0)	0
P24	Family	(0.51, 0)	0.12
	Friends	(0.40, 0)	0.08
	Acquaintances	(0.38, 0)	0
	Strangers	(0.02, 0)	0
P25	Family	(0.03, 0)	0
	Friends	(0, 0)	0
	Acquaintances	(0.30, 0)	0
	Strangers	(0, 0)	0
P26	Family	(0.36, 0)	0.06
	Friends	(0.38, 0)	0
	Acquaintances	(0.75, 0)	0.12
	Strangers	(0.19, 0)	0
P28	Family	(0.62,0)	0.10
	Friends	(0.57,0)	0.04
	Acquaintances	(0.27,0)	0
	Strangers	(0.11,0)	0
P29	Family	(0.27,0)	0
	Friends	(0.39,0)	0
	Acquaintances	(0.84,0)	0
	Strangers	(0.25,0)	0
P31	Family	(0.34,0)	0
	Friends	(0.72,0)	0.09
	Acquaintances	(0.25,0)	0
	Strangers	(0.09,0)	0
P32	Family	(0.75,0)	0.02
	Friends	(0.28,0)	0
	Acquaintances	(0.90,0)	0
	Strangers	(0.05,0)	0

7.6.5 Changes awareness

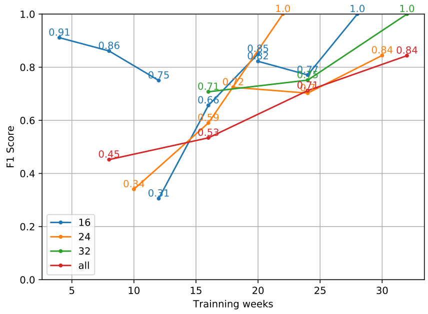
Rather than giving a specific number of ratings, changes could indicate the direction of the ratings and could immediately notify participants' deviation from their situation, which could be more evident than rating numbers. Therefore, change awareness is another aspect of monitoring the social interactions of participants. As a result, the aim of our model became the probability of social ratings changes under the state of features changes. Naive Bayes could learn each feature's conditional probability from existing data and predict the most likely outcome for a new case. It means a probability model could be built on the shift condition of features to predict the changes in social ratings. Different weights of communication channels given by participants when they completed social ratings could also be reflected by these probabilities. Thus, Naive Bayes is appropriate for our purpose. In practice, the Naive Bayes model faces a similar case as linear regression, as different measures could quantify social activities among each communication channel, so only one of them was picked in a single model. It also enables the independence of each feature, as Naive Bayes requires it; therefore, the whole model could be represented by the following equation:

$$P(Social\,ratings\,changes) = P(Face-to-face\binom{Variables}{1}|Social\,ratings\,changes)*$$
$$P(Call\,changes\binom{Variables}{1}|Social\,ratings\,changes)*$$
$$P(Messages\,changes\binom{Variables}{1}|Social\,ratings\,changes)*$$
$$P(Social\,media\,changes\binom{Variables}{1}|Social\,ratings\,changes)$$

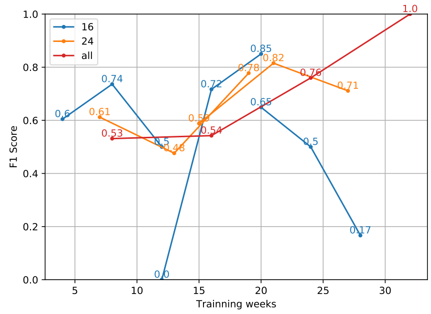
Under this strategy, features and ratings were further processed to generate differences. The numerical difference between this week and the previous week was calculated, and it was then divided by the previous week's value to obtain the percentage of change. We set 10% as the change threshold because it could denote a one-point change out of a total of 10 points in social ratings. That is to say, if the feature increased by more than 10%, it was treated as an increment and vice versa for a decrease of more than 10%. Other changes within 10% were regarded as no change.

For Naive Bayes, all ratings were transformed into three labels: increase, decrease and stable. Performance was assessed according to if the prediction was consistent with the actual condition. An overall score was termed an F1 score and was calculated from precision and recall. We used the weighted average to generate these indexes as a multiple labels model, meaning that precision, recall and the F1 score were calculated by each label. The label to be calculated was treated as positive, and other labels were treated as negative. Precision was the percentage of correctly predicted positive observations to predicted positive observations, which indicated a relevant instances ratio among the retrieved instances. The percentage of correctly predicted positive observations to all positive observations was reflected by recall, which measured the ratio of retrieved instances that were relevant. The F1 score combined the two measures as the weighted average of precision and recall.

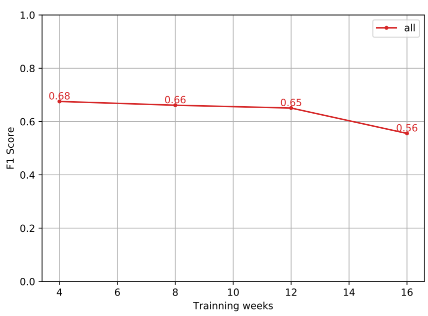
We followed the same training and testing process as the linear regression model to examine the Naive Bayes model. The weeks included, part separation and train/test procedure were identical, and the only difference was the total number of weeks of all participants minus one as there was no previous week for the first week to calculate the change. The results are shown in Figure 7.5. As a novel model, the comparison target was the random guess. As we had three labels (increase, decrease and stable), our model outperformed the random guess if the F1 score was larger than 1/3. As can be seen from the figure, the majority of results are larger than 0.33, which indicates our model surpasses random guesses. Generally, the F1 score still fluctuated with the increasing number of training weeks, and subsequent changes caused by the number of training weeks still varied. Nevertheless, unlike the linear regression model, the F1 score was always above 0.33 in all data parts except for P28. This could indicate that the Naive Bayes model could adapt to more extended social patterns than linear regression. For PDQ-8, we considered the minimal clinically important difference (MCID) as the change that needed attention. From a previous study [97], -5.94 and +4.91 points were the MCIDs for detecting improvement and worsening, respectively. However, the MCID did not always happen to all of our participants, as P28 and P31 had no MCID at all. The maximum number of changes appeared in P24, with only four MCIDs in a total of 34 weeks. Therefore, we did not include PDQ-8 in the Naive Bayes model. A similar situation happened in the types of contacts ratings, as there was a lack of variation in ratings from all participants. The importance of changes in a single type of contact was not significant as overall social ratings changed, so they were not included in the change awareness model.



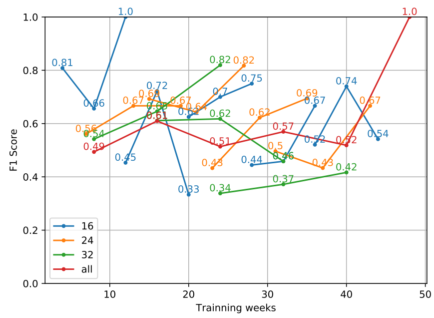
(a) P23



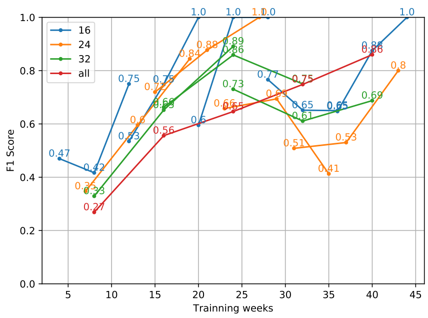
(b) P24



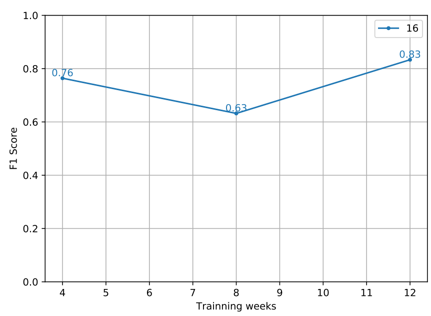
(c) P25



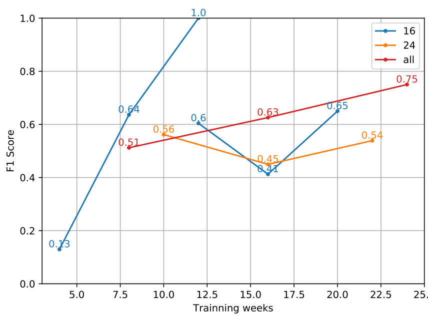
(d) P26



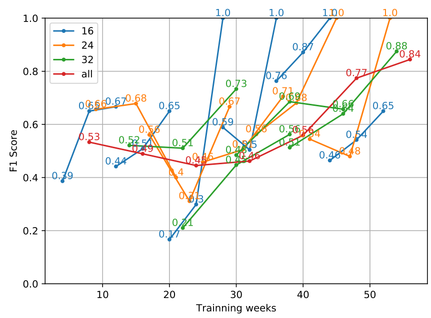
(e) P28



(f) P29



(g) P31



(h) P32

Figure 7.5: F1 score of the Naive Bayes model on overall social rating changes by every participant

7.7 Discussion

In this study, we completed a year-long longitudinal experiment with smartphones to observe PD patients' social behaviour. Twenty-three features were extracted from more than 10 million raw smartphone data to build personalised models of their social ratings. Different splits of every participant's complete dataset were tested to examine the performance of the models. The results demonstrate that we successfully established models to reflect the participants' extent of weekly social activities from both numeric values to change of direction. The values of standardised QoL ratings were also predicted proficiently from the model. The results underline the feasibility of our smartphone sensing approach in monitoring PD participants' weekly social interaction levels and QoL from extracted smartphone features. It has the potential to replace subjective scales and be regarded as a standardised objective monitoring tool in the future. Furthermore, the highest evaluation scores, model performance in different parts of data and the influence of increasing the number of training weeks are distinctive in every participant. This indicates the complexity of social behaviours, as each participant has their own social patterns and possibly their own PD progression path. Many factors influence social behaviour and disease progression, and it is impossible to control all of them when conducting a cohesive study. Social withdrawal in PD is a complicated phenomenon that should be understood on an individual basis.

For the number of times our models outperformed the benchmark (i.e. just assigning mean for the multiple linear regression model and random guess for Naive Bayes model), the Naive Bayes model was more successful than the multiple linear regression model, which was evident when the whole dataset was used to establish the model. For example, the R-squared of P31's multiple linear regression model of overall social ratings across all of the data was always 0. However, the minimum F1 score of P31's Naive Bayes model of overall social ratings was 0.51. This could be because direction prediction reduced the complexity of the model. For numerical prediction, the model had to learn the features for all 10 social ratings. In particular, some ratings that exist in future did not exist in the past, and it caused difficulties for the model to predict from non-existing ratings. This situation is better for direction prediction because the ground truth is simplified. Only three directions were considered, and almost all of these directions existed in every model. In general, researchers can choose different models according to the aim of their studies. If detailed levels of social activities are not required, direction prediction can generate more reliable results. Researchers can also set their own changes threshold to generate tailored change features.

Moreover, before the model was constructed, we expected its performance to improve with the number of training weeks because the model should be more robust with extensive information. However, the results show that the evaluation scores fluctuated with increasing training weeks. This could be explained by the lack of variation in self-report ratings in a certain period. For example, P24 gave nine overall social ratings from week six to week 12, and if the model had been constructed on this data part, the model would tend to give the test weeks nine ratings. Alternatively, if this data part became a significant part of the model, it would impact the generalisability of the prediction results. Therefore, choosing a reasonable time to establish the model significantly impacts model performance. Additionally, our observation period was during the COVID-19 pandemic, which severely changed participants' social lives as they had to stay at home, keep socially distanced and eliminate face-to-face interactions. Their social habits had to be changed due to external forces; therefore, the social patterns learnt pre-pandemic were unsuitable during the pandemic. This could also explain why the R-squared of the linear regression model on the whole data part was less than 0 in some instances.

The results also demonstrate that both overall social ratings and PDQ-8 models achieved reasonable evaluation scores. Furthermore, the combination of features in the model could also reveal the relationship between social activities and QoL level. However, it was found that all combinations of features achieving the highest evaluation score were different in every participant's data part, which suggests that the models constructed for PDQ and overall social ratings were totally different. Nevertheless, theoretically, as PD progression causes social withdrawal, some association between PDQ and overall social ratings is expected, so we conducted correlation analysis on the weekly social ratings and PDQ-8 for each participant. It was found that no coefficient was bigger than 0.6, which indicates there is probably no strong connection between PD progression and social behaviour. During the experiment, we also conducted comprehensive clinical and psychological Parkinson's and social withdrawal scales every two months, including the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [76], and Parkinson's Disease Questionnaire (PDQ-39) [102] and modified social withdrawal scale [185]. We conducted further correlation analysis on the social withdrawal scale and PDQ-39/MDS-UPDRS to explore the relationship between PD progression and social withdrawal from all questionnaire perspectives. Four participants had significant correlations for social withdrawal and PDQ-39 (P23: r=0.829, P<0.05, P26: r=0.745, P<0.1, P28: r=-0.891, P<0.05, P31: r=0.914, P<0.05) and three for social withdrawal and MDS-UPDRS (P26: r=0.918, P<0.01, P28: r=-0.916, P<0.05, P32: r=0.985, P<0.01). From the questionnaires, this analysis suggests that not all participants' social withdrawal was closely connected with PD progression. This result also confirms the findings from the feature combination difference that there is a possibility that PD progression and social withdrawal are independent.

7.8 Limitations and Future Work

Although our model successfully learnt individual social behaviour at a granular level, it still suffers from potential issues that can be dealt with in future work.

PD is a long-term disease, and we initiated a year-long longitudinal study to observe participants' social behaviours. However, the duration of the experiment may not be enough to observe significant changes, which is valuable in determining the impacts on participants. As stated in the results, MCIDs in PDQ-8 does not appear in two out of eight participants. Even for the participants with the highest number of MCIDs, it only appears four times in the observations over the entire year. The limited number of MCIDs makes it impossible to construct the models for these changes, so extended experiments are necessary to provide more chances to observe these MCIDs, and significant smartphone features can be detected when these MCIDs happen.

Even successful prediction was shown for the weekly social ratings, but this does not mean that our model covers all social interactions of participants. Participant interviews indicated that they still had other means of communication, and landline phones made up a significant part of their calls. Therefore, their usage of smartphones could influence the general performance of the model, which is an intriguing question to be studied. For participants who mostly interacted socially on smartphones, the performance of our model should be better because fewer interactions will be missed. Moreover, to establish a comprehensive observation of participants, more monitoring methods should be implemented to cover all possible communication channels.

Compared to the overall model, the performance of models for different types of contacts was not significant. One of the critical reasons for this was that the technique for differentiating types of contacts is still at an early stage. To the best of our knowledge, a state-of-the-art study utilised the frequency of each communication channel to differentiate types of contacts, but this technique is not clinically validated. Real-life is more complex than the preliminary results of these studies, and contacts with the highest frequency do not always mean that participants' relationships are with families and close friends. Further work can explore more reasonable unobtrusive approaches to decide types of contacts so that an in-depth understanding of each social activity can be achieved.

7.9 Conclusion

From the basis that PD progression causes social withdrawal and the treatment aim is to improve PD patients' QoL, we initiated a longitudinal experiment to study the social behaviour of PD patients using smartphones. By extracting features from raw data and considering the nature of social contact, we successfully applied the social behaviour model to predict social interaction levels and QoL based on collected smartphone data. Our two models successfully reflected the levels of social interaction and QoL from different perspectives in the granularity of a week.

Moreover, our method adapted to each participant and made predictions for that particular person. This meets the idiosyncratic nature of PD progression and individualised social patterns. It also enables personal understanding of social withdrawal and PD, which provides foundations for personalised care and precision medicine. Furthermore, our model demonstrates the potential to reconstruct social interactions through different types of contacts. With the development of a technique to distinguish types of contacts, our model could perform better for overall social interaction levels and PDQ.

To the best of our knowledge, this is the first study to use smartphones to establish personal weekly social behaviour models for PD patients. The models provide significantly detailed tracking of social behaviours of PD patients from an objective perspective and could benefit social behaviour studies in PD and broader communities where the social impact of patients needs attention.

Chapter 8

Conclusions and future work

In this work, we explored social withdrawal via smartphones in people with Parkinson's. This work was built on two pillars: 1) PD induces a series of emotional and communicative changes in patients, disrupting their social functions [178]; 2) an unobtrusive and personalised digital-phenotyping technology can be used to monitor the behaviour of PD patients [94]. The data from smartphones served as the foundation for a digital phenotyping method that intended to extract social-behaviour occurrences and features. As a long-term neurodegenerative disease, the social withdrawal caused by PD should be monitored in a longitudinal manner. Therefore, we conducted a year-long longitudinal study to observe social withdrawal in PD patients. All these conclusions revolve around this experiment.

Primarily, we conducted a systematic review of passive smartphone social sensing to extract the technical basis of this technology (Chapter 3). The systematic review comprehensively explored past studies utilising passive smartphone social sensing. The paradigm of these studies was discovered. Fundamentally, a monitoring application was running on participants' smartphones throughout the whole experiment. Other validation measures, such as questionnaires or tasks, were conducted twice or several times at the beginning, end or in the middle. These measures were then analysed with features interpreted from raw collected data to examine the study's hypothesis. The fundamental resolution included gathering social interactions on smartphones, and inferring social interactions outside smartphones was also presented. To reconstruct the social behaviours of participants, calls, messages and social media usage were three major sources of social interactions on smartphones. Bluetooth microphones could be used to infer social activities outside the smartphone, such as face-to-face conversations. Although passive smartphone social sensing has limitations, such as accuracy issues and privacy challenges, the reviewed studies showed its feasibility in observing social behaviours from an objective perspective. Our experiment followed the methodological results of the reviewed literature. In the following chapters on PD patients, we applied passive smartphone social sensing in the wild.

An application was chosen to install on participants' smartphones to collect designated sensor data 24 hours, seven days per week. A set of clinical/psychological scales considering PD progression, QoL, social withdrawal and related issues were planned to be applied every two months. Specially designed diaries were prepared to ask participants to self-report their weekly social extent and QoL. After ethical approval, participants' recruitment started in September 2019 and ended in March 2020.

Unfortunately, COVID-19 became a worldwide health issue since then. The UK government introduced a series of policies to reduce the transmission of diseases. Although COVID-19 severely interfered with our research, Chapter 4 and Chapter 5 regarded it as an opportunity to examine if our monitoring technology can detect these impacts. In Chapter 4, from cross-participant collaboration, our approaches demonstrated the general capability of reflecting the social behaviour changes of participants. Features and smartphone-data perspectives were created from both policies. Home locations were identified from GPS, and it appeared that times of leaving the house decreased significantly. The Bluetooth inferred that participants had fewer face-to-face contacts compared with the time before the pandemic. All participants' data showed a similar declining trend, indicating they all obeyed the government rules to minimise travel and social contact with others. Chapter 5 focuses on individual social adjustment towards COVID-19–related policies. Although the pandemic and corresponding lockdown severely impacted the basis of our observation, people could still have social interaction as they wished. Our approach still demonstrates its feasibility of detecting participants' social behaviour changes. Through semi-structured interviews with participants, we confirmed that collected smartphone data captured their personalised reactions towards the policies introduced by COVID-19. They adapted novel call and message patterns and left home for different purposes. Furthermore, the preliminary smartphone data also exhibited levels of coherence with participants' diary ratings. This suggested that the features created and models constructed reasonably reflected participants' daily social behaviours. These findings and observation approaches could be used for delivering specialised care and future policy-making. They confirm that our approach can present the trend of all participants and reveal alterations at the individual level.

The longitudinal observation ended in June 2021, and eight participants finished the whole year of data collection. More than ten million raw smartphone-data entries and three thousand days of diaries were collected in total. Psychological/clinical scales were retrieved from each participant at least five times. These psychological/clinical scales included Parkinson's progression, QoL, social withdrawal, cognition, stigma, depression, empathy and apathy. Combined with weekly social and QoL ratings in diaries, they were treated as ground truth in our analysis. As mentioned in the introduction chapter, two research aims of our study were 1) make connections between social behaviour collected by digital phenotyping and the clinical/psychological ground truth of Parkinson's and 2) personalise social-behaviour tracking in a more granular manner. Therefore, we conducted two levels of analysis on smartphone data and ground truth. The first level was towards those psychological/clinical scales, and the second was towards weekly ratings, since they are more granular than scales. These two levels of analysis form Chapter 6 and Chapter 7.

The smartphone features were constructed according to communication channels and location factors. Different measures, including length, time, unique contact of these channels and semantic understanding of locations, were all considered. A conceptual model involving the intimacy of contacts for in-depth measuring of the social withdrawal in PD patients was also proposed. Then, we conducted the correlation analysis between the smartphone features and the summary index of these scales. The full results are discussed in Chapter 6. At least one feature was found to have significant correlations with clinical/psychological scales for each participant. As the scale most relevant to social behaviour, the significant correlation between the social withdrawal scale and the smartphone feature was found in every participant. It demonstrated the potential of the smartphone as an objective social-behaviour measurement. Moreover, smartphone data showed that there tends to be reduced social activity when situations are worse. They provided evidence that Parkinson's progression could have a positive relationship with social withdrawal.

After the standardised scales correlation, we turned to personal-model construction based on more granular diaries. A feature selection and combination mechanism were applied to aggregate all channels of communication and related factors. Our model considered numerical and direction predictions, so both level and changes could be learnt. By applying multiple linear regression and Naive Bayes, our model obtained at least 0.6 R-squared in numerical prediction and 0.6 F1 score in direction projection of the highest value of all participants. This was significantly better than just assigning mean or random guesses, which illustrates the effectiveness of our approach. The results also exhibit the smartphone's potential for monitoring social behaviours and QoL weekly. For each participant, the significant correlated features and feature combinations that achieved the highest scores were different. All these results also confirm the personal difference between participants. It is necessary to observe participants' social behaviour individually.

8.1 Main findings

- **Our smartphone sensing approach is feasible and effective as a digital phenotyping method for monitoring social behaviour. It has the potential for an in-depth understanding of social activities.** Overall, our approach demonstrated the feasibility of a unique method of digital phenotyping to observe people's social behaviour. In both COVID-19 and Parkinson's studies, its effectiveness was demonstrated thoroughly. By choosing appropriate features, participants' conformance with policies and the social changes due to the impact of COVID-19 were opportunely reflected by the smartphones. Results of the longitudinal study also demonstrated the ability of our smartphone sensing approach to reveal social behaviour from an objective perspective. We also established a social behaviour model considering communication channels and contact types in terms of Parkinson's. The model's successful results demonstrate our approach's ability to construct an in-depth understanding of social activities, which takes digital phenotyping a step further. Every smartphone's data is exclusively generated by the participant, which also enables our approach to understand the social behaviour of particular participants. Moreover, raw data from a smartphone can be interpreted in a variety of ways, which creates extensive possibilities for researchers to utilise them. Therefore, digital phenotyping is promising as a standardised measurement for social behaviour, and its unobtrusiveness can benefit novel clinical/psychological research.

- **Possible positive correlations between PD progression and social withdrawal were found, but it is a complex phenomenon that needs further investigation.** Previous interview- or questionnaire-based studies suggested that PD causes social dysfunction in PD patients, inducing social withdrawal. To the best of our knowledge, this phenomenon has never been studied in a long-term manner. Our experiment is the first to explore social withdrawal in PD objectively and longitudinally. Significant correlations were found between clinical/psychological scales and smartphone features. For all participants, most of the results indicated that social withdrawal and related factors became more severe as PD worsened. However, there were some results that did not meet our expectations. Particular participants tended to have more of certain social interactions when situations were worse. These results provide evidence of possible positive correlations between PD progression and social withdrawal but also reveal the complexity of this phenomenon. Naturally, various factors influence levels of social interaction; how PD alone impacts patients requires further investigation.

- **Adaptability and personalisation are essential in social behaviour research.** We discussed the importance of individualised monitoring in every chapter involved in the experiment. The results of different individuals in all studies indicated that their situations were diverse. Every participant had different reactions towards the impacts of COVID-19. The significantly correlated smartphone features of the clinical/psychological scales were also different in each individual. Accordingly, we were able to develop an adaptive technique capable of reacting to social dynamics, generating a personalised model from multiple data sources and producing an overfitted collection of social behaviour models. Moreover, feature combinations of these models achieving the highest evaluation metric were different, which confirmed the individuality of participants. Preliminary results from the models considering different types of contact also exhibited their potential for in-depth understanding of the social behaviour of a particular individual. In summary, as behaviour is influenced by various factors, adaptability and personalisation are prerequisite considerations of social behaviour research.

8.2 Limitations

Our studies are based on collecting social behaviour data from smartphones. They are a reliable source for social interactions since all data were recorded directly on the smartphones. Social interactions outside of smartphones, such as face-to-face conversations, were inferred from Bluetooth and microphone. The reliability of this inference, as well as the accuracy, is unknown. This issue also applies to location determination. Although parameters were learnt from the previous experience, it is not absolutely certain that participants really went to that number of places. Moreover, we relied on previous research [84] to differentiate types of contact. But, its concept is not universally validated. The results do not indicate the type of relationships between the contacts and participants. We do note the challenges of this approach, but it is still satisfactory for our study because we focus on the general trend rather than specific behaviour. But, with the increasing performance of every inference, the available results could have improved accuracy.

As a smartphone-monitoring study, we comprehensively considered all possible social features the smartphone could capture. Although people spend significant time on smartphones for social interactions, and social interactions outside smartphones can be inferred, it is still not their entire social lives. Social interactions could happen on other devices, such as laptops, tablets or landlines. In addition, our approach is limited to publicly available smartphone data. This means that these data are retrievable from the operating system of the smartphones, and the privacy of participants is considered. However, this approach is not necessarily the best at deriving specific features of social activity. Theoretically, smartphones could provide more critical data, which involves communication details such as content or emotion. These data could establish an extensive understanding of participants' social behaviour and build more precise models of personal contact. The intimacy of each communication could also be differentiated with more direct data sources. However, the monitoring application needs further implementation to acquire these data, and there will be more privacy concerns from participants.

As an exploratory study, the generalisability of the results has not been examined. Our study did not specifically attribute social withdrawal to PD progression. In reality, PD might not be the only cause of social withdrawal nor the determination of social interaction level. For example, COVID-19 and related constantly changing policies undoubtedly impact participants' social lives. Personality is also a significant factor for sociality. Extroverts and introverts tend to have different social behavioural patterns [104]. In this study, we only observe the social courses of PD patients, but comparing it with other people with similar demographics would be beneficial. The control group without PD could be introduced at the beginning to enhance the validity of the study. Direct contrasts could be made between the control group and the experiment group, so the uniqueness of social withdrawal in PD would be extracted. The reasons behind social withdrawal could also be acknowledged by exploring the difference between the experiment and the control group. Moreover, there are different factors such as medication, physical activity and nutrition that could influence the social situation of PD patients. The stages of PD and the treatment participants experienced could also affect the extent of social withdrawal [178]. As far as personalised monitoring, we focused on individual social behaviour and disease progression. Although our approach can be used for studying personal social behaviour, our results cannot be expanded to a broader population, since the baseline was established according to these particular participants. Therefore, how PD alone causes social withdrawal in general needs further investigation. By introducing control groups, the reasons behind social withdrawal could be acknowledged by exploring the different situations between the experiment and the control group. Even more, researchers could manipulate particular variables like medication or treatment to examine their impact on the social lives of PD patients. Stages of PD and dose of medication can be considered when designing the experiment. Overall, the control group can contribute to a deeper and more specific understanding of social withdrawal in PD.

In addition, Bluetooth signals can not be fully treated as a person. The accuracy of using Bluetooth for proximity and social interaction detection needs further research. Firstly, not only smartphones but also other digital devices are equipped with Bluetooth. For example, a person could own a smartwatch, tablet or smartphone, and these devices are around at the same time. So the detected signal could be duplicated if treated as a person. Secondly, the latest Android system has adopted a privacy strategy that the unique identifier of Bluetooth – MAC address could be different at every scan. It increases the uncertainty that different Bluetooth signals may not represent unique contact but the same person. Moreover, there are additional factors that may influence the transmission of Bluetooth. Factors like indoor or outdoor, smartphone position, and weather conditions will affect the quality of the scan and receive. Particularly, Bluetooth signals could be noisy in moving conditions because received signals could have gone when they are recorded. Therefore, it's significant to investigate how to reasonably utilise Bluetooth to infer proximity and face-to-face interactions.

Although we innovated an unobtrusive method to monitor people's daily lives, the ground truth we relied on is still self-report diaries and scales. Even if we asked participants to behave normally and have activities as they usually do, the Hawthorne effect introduced by these measures could not be eliminated. Participants may still feel monitored during the two-month visit when questions were asked, and the diary could remind them they are involved in a study of social interactions. When participants see the icon of the installed monitoring application, they may also be aware that their interactions on smartphones are recorded. These reminders could always make participants feel they are being observed, so they may change their behaviour as it is being evaluated. These potential weaknesses impact the validity of the whole study and need additional consideration when conducting similar experiments.

Furthermore, the ground truth we relied on was self-report or scales conducted by humans, which suffer the same potential issues as the questionnaires. Their results are influenced by selective memory, telescoping, attribution and exaggeration [80]. To alleviate some of these effects, we introduced diaries to participants, so they could record their situations instantly. But, the natural deficits of self-report cannot be fully compromised. Moreover, these scales are the only clinically or psychologically verified equipment before new valid and objective tools are invented. To reach the significance of the measurement, we do not have other options that have the same legitimacy.

8.3 Future work

Several proposals for further study have been mentioned in earlier chapters. In this section, we summarise those potential prospects in the context of the bigger picture:

Rational sensor choice: From Chapter 3, energy consumption was the primary consideration for sensing strategy out of all the reviewed digital-phenotyping studies. However, does this limit account for the entire rationale for the researchers' decision? Is the result always better with the highest sensor rate if energy consumption is not an issue? The frequency of sensor data may have an impact on these outcomes. Different resolutions can either improve performance or increase interference for data processing. A better result with less energy consumption might be achieved using a more efficient sensor frequency. Reasonable sensor and parameter choice is the crucial question in digital phenotyping. The rationale for choosing an optimal sensor rate should be considered from the purpose of the study. Further investigation could be done on the impact of different sensor rates. A nontrivial method for developing the sensing strategy could be to start with the study's purpose. Experiments with various sample rates and sensor combinations could be undertaken to investigate which option is most effective in reaching the study's target. Despite the existence of research attempting to resolve these concerns, such as [131], which studied the required Bluetooth signal strength for inferring face-to-face proximity in various settings, there are still many unanswered issues in this sector. For example, how frequently should the Bluetooth scan be activated to recover a particular proportion of the face-to-face contacts of participants, and how does the frequency of Bluetooth scans influence the accuracy of face-to-face recovery? These investigations will also provide clinical value to health-related research and create a strong foundation for digital-phenotyping sensor usage.

From correlation to causation: Although the feasibility of digital phenotyping has been confirmed in our studies, from Chapter 3, the most published results are the correlation coefficient and performance of the machine learning algorithm applied without a detailed explanation of the cause. Strong correlations or classifications do not always imply valid measures, particularly when demographics and sample sizes are limited. These limitations may jeopardise the digital-phenotyping results' generalisability and effectiveness. For example, a study of personality using digital phenotyping reported that their findings were not consistent with previous studies [208]. The different types of data used explained this inconsistency. On the basis of existing digital-phenotyping results, more theoretical research or cross-population trials should be considered. Researchers could treat smartphone features like questionnaire items, so they can be evaluated and interpreted in the clinical or psychological method. Variables derived from smartphone data could be unified by time. The number of calls, for example, could be gathered weekly or monthly. Therefore, comparison between digital-phenotyping studies will become possible. Then, instead of being a promising field, digital phenotyping may become the standardised measurement.

Clinical PD research: Because it is a long-term disease and the major patient population is the elderly, Parkinson's impact on social behaviours coexists with many other factors, such as age, mobility and personality. Specialised experiment designs and dedicated data processing techniques are necessary to study PD-specific social withdrawal. Moreover, smartphone monitoring alone cannot cover the entire social lives of participants. To achieve thorough social behaviour monitoring, all equipment for social interaction, such as landline phones, tablets and handwritten mail, should be considered. Apart from social behaviour, there are plenty of novel research topics, such as cognition, attention, memory and mobility, for PD behavioural studies in which digital phenotyping can be applied. The achievement of these studies can give a complete picture of the QoL monitoring of PD patients. The future work proposed by the above two paragraphs also applies to PD digital-phenotyping research. With reasonable sensor-choice and clinical-causation studies, smartphones have the potential to replace existing questionnaires as a standardised PD clinical assessment. So, the granular and comprehensive monitoring of PD patients can be achieved by a combination of obtrusive and unobtrusive methods. With clinical validity, digital phenotyping can provide carers and clinicians with immediate PD progression and QoL information. Then, better care and treatment can be provided to PD patients to improve their QoL.

Bibliography

[1] S. Abdullah, M. Matthews, E. L. Murnane, G. Gay, and T. Choudhury. Towards circadian computing: " early to bed and early to rise" makes some of us unhealthy and sleep deprived. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pages 673–684, 2014.

[2] L. Abou, J. Peters, E. Wong, R. Akers, M. S. Dossou, J. J. Sosnoff, and L. A. Rice. Gait and balance assessments using smartphone applications in parkinson's disease: A systematic review. *Journal of medical systems*, 45(9):1–20, 2021.

[3] N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland. The social fmri: measuring, understanding, and designing social mechanisms in the real world. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 445–454. ACM, 2011.

[4] H. Allcott, L. Boxell, J. Conway, M. Gentzkow, M. Thaler, and D. Y. Yang. Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *NBER Working Paper*, (w26946), 2020.

[5] M. Andersen. Early evidence on social distancing in response to covid-19 in the united states. *Available at SSRN 3569368*, 2020.

[6] P. Andriotis, A. Takasu, and T. Tryfonas. Smartphone message sentiment analysis. In *IFIP International Conference on Digital Forensics*, pages 253–265. Springer, 2014.

[7] S. Aral, E. Brynjolfsson, and M. Van Alstyne. Information, technology and information worker productivity: Task level evidence, 2007.

[8] S. Arora, V. Venkataraman, A. Zhan, S. Donohue, K. M. Biglan, E. R. Dorsey, and M. A. Little. Detecting and monitoring the symptoms of parkinson's disease

using smartphones: a pilot study. *Parkinsonism & related disorders*, 21(6):650–653, 2015.

[9] A. Barrat, C. Cattuto, V. Colizza, J.-F. Pinton, W. V. d. Broeck, and A. Vespignani. High resolution dynamical mapping of social interactions with active rfid. *arXiv preprint arXiv:0811.4170*, 2008.

[10] D. Barua, J. Kay, and C. Paris. Viewing and controlling personal sensor data: what do users want? In *International Conference on Persuasive Technology*, pages 15–26. Springer, 2013.

[11] G. F. Bati and V. K. Singh. “trust us”: Mobile phone use patterns can predict individual trust propensity. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 330. ACM, 2018.

[12] G. Bauer and P. Lukowicz. Can smartphones detect stress-related changes in the behaviour of individuals? In *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, pages 423–426. IEEE, 2012.

[13] R. F. Baumeister and M. R. Leary. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, 117(3):497, 1995.

[14] À. Bayés, A. Samá, A. Prats, C. Pérez-López, M. Crespo-Maraver, J. M. Moreno, S. Alcaine, A. Rodriguez-Molinero, B. Mestre, P. Quispe, et al. A “holter” for parkinson’s disease: Validation of the ability to detect on-off states using the rempark system. *Gait & posture*, 59:1–6, 2018.

[15] N. K. Baym, Y. B. Zhang, and M.-C. Lin. Social interactions across media: Interpersonal communication on the internet, telephone and face-to-face. *New Media & Society*, 6(3):299–318, 2004.

[16] R. Beckwith and S. Mainwaring. Privacy: Personal information, threats, and technologies. In *Proceedings. 2005 International Symposium on Technology and Society, 2005. Weapons and Wires: Prevention and Safety in a Time of Fear. ISTAS 2005.*, pages 9–16. IEEE, 2005.

[17] D. Ben-Zeev, E. A. Scherer, R. Wang, H. Xie, and A. T. Campbell. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric rehabilitation journal*, 38(3):218, 2015.

[18] J. D. Berry, S. Paganoni, K. Carlson, K. Burke, H. Weber, P. Staples, J. Salinas, J. Chan, J. R. Green, K. Connaghan, et al. Design and results of a smartphone-based digital phenotyping study to quantify als progression. *Annals of clinical and translational neurology*, 6(5):873–881, 2019.

[19] R. Bhidayasiri and Z. Mari. Digital phenotyping in parkinson's disease: Empowering neurologists for measurement-based care. *Parkinsonism & Related Disorders*, 80:35–40, 2020.

[20] T. Biloborodova, I. Skarga-Bandurova, O. Berezhnyi, M. Nesterov, and I. Skarha-Bandurov. Multimodal smartphone-based system for long-term monitoring of patients with parkinson's disease. In *International Conference on Information Technology & Systems*, pages 626–636. Springer, 2020.

[21] B. Bolluk, E. T. Özel-Kizil, M. C. Akbostanci, and E. C. Atbasoglu. Social anxiety in patients with parkinson's disease. *The Journal of neuropsychiatry and clinical neurosciences*, 22(4):390–394, 2010.

[22] B. M. Bot, C. Suver, E. C. Neto, M. Kellen, A. Klein, C. Bare, M. Doerr, A. Pratap, J. Wilbanks, E. Dorsey, et al. The mpower study, parkinson disease mobile data collected using researchkit. *Scientific data*, 3(1):1–9, 2016.

[23] M. Boukhechba, Y. Huang, P. Chow, K. Fua, B. A. Teachman, and L. E. Barnes. Monitoring social anxiety from mobility and communication patterns. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 749–753, 2017.

[24] A. Boytsov, A. Zaslavsky, and Z. Abdallah. Where have you been? using location clustering and context awareness to understand places of interest. In *Internet of Things, Smart Spaces, and Next Generation Networking*, pages 51–62. Springer, 2012.

[25] K. C. Breen and G. Drutyte. Non-motor symptoms of parkinson's disease: the patient's perspective. *Journal of neural transmission*, 120(4):531–535, 2013.

[26] E. Brietzke, E. R. Hawken, M. Idzikowski, J. Pong, S. H. Kennedy, and C. N. Soares. Integrating digital phenotyping in clinical characterization of individuals with mood disorders. *Neuroscience & Biobehavioral Reviews*, 104:223–230, 2019.

[27] G. Briscese, N. Lacetera, M. Macis, and M. Tonin. Compliance with covid-19 social-distancing measures in italy: the role of expectations and duration. Technical report, National Bureau of Economic Research, 2020.

[28] B. Buck, K. A. Hallgren, E. Scherer, R. Brian, R. Wang, W. Wang, A. Campbell, T. Choudhury, M. Hauser, J. M. Kane, et al. Capturing behavioral indicators of persecutory ideation using mobile technology. *Journal of psychiatric research*, 116:112–117, 2019.

[29] B. Buck, E. Scherer, R. Brian, R. Wang, W. Wang, A. Campbell, T. Choudhury, M. Hauser, J. M. Kane, and D. Ben-Zeev. Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. *Schizophrenia research*, 208:167–172, 2019.

[30] J. T. Cacioppo, L. C. Hawkley, and G. G. Berntson. The anatomy of loneliness. *Current directions in psychological science*, 12(3):71–74, 2003.

[31] M. D. Carlson and R. S. Morrison. Study design, precision, and validity in observational studies. *Journal of palliative medicine*, 12(1):77–82, 2009.

[32] S. Centellegher, M. De Nadai, M. Caraviello, C. Leonardi, M. Vescovi, Y. Ramadian, N. Oliver, F. Pianesi, A. Pentland, F. Antonelli, et al. The mobile territorial lab: a multilayered and dynamic view on parents' daily lives. *EPJ Data Science*, 5(1):3, 2016.

[33] L. M. Chatters, H. O. Taylor, E. J. Nicklett, and R. J. Taylor. Correlates of objective social isolation from family and friends among older adults. In *Healthcare*, volume 6, page 24. Multidisciplinary Digital Publishing Institute, 2018.

[34] K. R. Chaudhuri, D. G. Healy, and A. H. Schapira. Non-motor symptoms of parkinson's disease: diagnosis and management. *The Lancet Neurology*, 5(3):235–245, 2006.

[35] G. Chittaranjan, J. Blom, and D. Gatica-Perez. Who's who with big-five: Analyzing and classifying personality traits with smartphones. In *2011 15th Annual international symposium on wearable computers*, pages 29–36. IEEE, 2011.

[36] G. Chittaranjan, J. Blom, and D. Gatica-Perez. Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, 17(3):433–450, 2013.

[37] Y. Chon, H. Shin, E. Talipov, and H. Cha. Evaluating mobility models for temporal prediction with high-granularity mobility data. In *2012 IEEE International Conference on Pervasive Computing and Communications*, pages 206–212. IEEE, 2012.

[38] D. Christin, A. Reinhardt, M. Hollick, and K. Trumpold. Exploring user preferences for privacy interfaces in mobile sensing applications. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, pages 1–10, 2012.

[39] M. Ciesielska, K. W. Boström, and M. Öhlander. Observation methods. In *Qualitative methodologies in organization studies*, pages 33–52. Springer, 2018.

[40] V. A. Coelho and A. M. Romão. The relation between social anxiety, social withdrawal and (cyber) bullying roles: A multilevel analysis. *Computers in Human Behavior*, 86:218–226, 2018.

[41] S. Coghlan, M. Cheong, and B. Coghlan. Tracking, tracing, trust: contemplating mitigating the impact of covid-19 through technological interventions. *interventions*, 213:6–8, 2020.

[42] A. S. Cohen, E. Schwartz, T. Le, T. Cowan, C. Cox, R. Tucker, P. Foltz, T. B. Holmlund, and B. Elvevåg. Validating digital phenotyping technologies for clinical use: the critical importance of "resolution". *World Psychiatry*, 19(1):114, 2020.

[43] V. P. Cornet and R. J. Holden. Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of biomedical informatics*, 77:120–132, 2018.

[44] C. E. Coyle and E. Dugan. Social isolation, loneliness and health among older adults. *Journal of aging and health*, 24(8):1346–1363, 2012.

[45] A. W. DaSilva, J. F. Huckins, R. Wang, W. Wang, D. D. Wagner, and A. T. Campbell. Correlates of stress in the college environment uncovered by the application of penalized generalized estimating equations to mobile sensing data. *JMIR mHealth and uHealth*, 7(3):e12084, 2019.

[46] M. H. Davis. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1):113, 1983.

[47] A. De Boer, W. Wijker, J. Speelman, and J. De Haes. Quality of life in patients with parkinson's disease: development of a questionnaire. *Journal of Neurology, Neurosurgery & Psychiatry*, 61(1):70–74, 1996.

[48] Y.-A. de Montjoye, J. Quoidbach, F. Robic, and A. S. Pentland. Predicting personality using novel mobile phone-based metrics. In *International conference on social computing, behavioral-cultural modeling, and prediction*, pages 48–55. Springer, 2013.

[49] S. Del Din, A. Godfrey, C. Mazzà, S. Lord, and L. Rochester. Free-living monitoring of parkinson's disease: Lessons from the field. *Movement Disorders*, 31(9):1293–1313, 2016.

[50] T. R. N. Desk. Smartphones become the central hub for individuals, 2016.

[51] A. Developers. Apple ios core motion, 2020.

[52] G. Developers. Google activity recognition api, 2020.

[53] T. M. Do and D. Gatica-Perez. Human interaction discovery in smartphone proximity networks. *Personal and Ubiquitous Computing*, 17(3):413–431, 2013.

[54] T. M. T. Do and D. Gatica-Perez. Groupus: Smartphone proximity data and human interaction type mining. In *2011 15th Annual International Symposium on Wearable Computers*, pages 21–28. IEEE, 2011.

[55] T. Donker, K. Petrie, J. Proudfoot, J. Clarke, M.-R. Birch, and H. Christensen. Smartphones for smarter delivery of mental health programs: a systematic review. *Journal of medical Internet research*, 15(11):e247, 2013.

[56] R. I. Dunbar. The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 6(5):178–190, 1998.

[57] N. Eagle and A. S. Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

[58] Editorials. The cooperative human. *Nature Human Behaviour*, 2018.

[59] T. Elmer, K. Chaitanya, P. Purwar, and C. Stadtfeld. The validity of rfid badges measuring face-to-face interactions. *Behavior research methods*, 51(5):2120–2138, 2019.

[60] P. Eskes, M. Spruit, S. Brinkkemper, J. Vorstman, and M. J. Kas. The sociability score: App-based social profiling from a healthcare perspective. *Computers in Human Behavior*, 59:39–48, 2016.

[61] A. J. Espay, P. Bonato, F. B. Nahab, W. Maetzler, J. M. Dean, J. Klucken, B. M. Eskofier, A. Merola, F. Horak, A. E. Lang, et al. Technology in parkinson's disease: challenges and opportunities. *Movement Disorders*, 31(9):1272–1282, 2016.

[62] D. Estrin. Small data, where n= me. *Communications of the ACM*, 57(4):32–34, 2014.

[63] A. A. Farhan, C. Yue, R. Morillo, S. Ware, J. Lu, J. Bi, J. Kamath, A. Russell, A. Bamis, and B. Wang. Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. In *2016 IEEE Wireless Health (WH)*, pages 1–8. IEEE, 2016.

[64] K. Farrahi and D. Gatica-Perez. Probabilistic mining of socio-geographic routines from mobile phone data. *IEEE Journal of Selected Topics in Signal Processing*, 4(4):746–755, 2010.

[65] M. Faurholt-Jepsen, M. Bauer, and L. V. Kessing. Smartphone-based objective monitoring in bipolar disorder: status and considerations. *International journal of bipolar disorders*, 6(1):1–7, 2018.

[66] M. Faurholt-Jepsen, J. Busk, H. órarinsdóttir, M. Frost, J. E. Bardram, M. Vinberg, and L. V. Kessing. Objective smartphone data as a potential diagnostic marker of bipolar disorder. *Australian & New Zealand Journal of Psychiatry*, 53(2):119–128, 2019.

[67] D. Ferreira, V. Kostakos, and A. K. Dey. Aware: mobile context instrumentation framework. *Frontiers in ICT*, 2:6, 2015.

[68] Finder.com. Mobile internet statistics, 2020.

[69] J. M. Fisher, N. Y. Hammerla, T. Ploetz, P. Andras, L. Rochester, and R. W. Walker. Unsupervised home monitoring of parkinson's disease motor symptoms using body-worn accelerometers. *Parkinsonism & related disorders*, 33:44–50, 2016.

[70] C. for Disease Control and Prevention. Mental health and coping during covid-19, Apr 2020.

[71] P. Foundation. Myths about parkinson's, 2021.

[72] Foursquare. The trusted location data and intelligence company, Apr 2020.

[73] J. Fox and B. McEwan. Distinguishing technologies for social interaction: The perceived social affordances of communication channels scale. *Communication Monographs*, 84(3):298–318, 2017.

[74] Y. Fukazawa, T. Ito, T. Okimura, Y. Yamashita, T. Maeda, and J. Ota. Predicting anxiety state using smartphone-based passive sensing. *Journal of biomedical informatics*, 93:103151, 2019.

[75] S. Gao, W. Li, L. J. Song, X. Zhang, M. Lin, and S. Lu. Personalitysensing: A multi-view multi-task learning approach for personality detection based on smartphone usage. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2862–2870, 2020.

[76] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel, et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society*, 23(15):2129–2170, 2008.

[77] GOV.UK. Coronavirus (covid-19): what you need to do, May 2020.

[78] T. Groceries. Tesco covid-19 updates, Apr 2020.

[79] T. W. Group. The world health organization quality of life assessment (whoqol): development and general psychometric properties. *Social science & medicine*, 46(12):1569–1585, 1998.

[80] R. M. Groves. *Survey errors and survey costs*, volume 536. John Wiley & Sons, 2004.

[81] B. K. Gultekin, B. Ozdilek, and E. E. Bestepe. Social phobia in parkinson's disease: Prevalence and risk factors. *Neuropsychiatric disease and treatment*, 10:829, 2014.

[82] S. D. Gunnery, B. Habermann, M. Saint-Hilaire, C. A. Thomas, and L. Tickle-Degnen. The relationship between the experience of hypomimia and social wellbeing in people with parkinson's disease and their care partners. *Journal of Parkinson's disease*, 6(3):625–630, 2016.

[83] B. Guo, Y. Liang, Z. Yu, M. Li, and X. Zhou. From mobile phone sensing to human geo-social behavior understanding. *Computational Intelligence*, 32(2):240–258, 2016.

[84] M. Gustarini. *Analysing smartphone users "inner-self": the perception of intimacy and smartphone usage changes*. PhD thesis, University of Geneva, 2016.

[85] M. Gustarini and K. Wac. Estimating people perception of intimacy in daily life from context data collected with their mobile phone. In *Mobile Data Challenge MDC 2012 (by Nokia) Workshop organized in conjunction with the 20th International Conference on Pervasive Computing 2012*, page 6, 2012.

[86] A. Haahr, M. Kirkevold, E. O. Hall, and K. Østergaard. Living with advanced parkinson's disease: a constant struggle with unpredictability. *Journal of advanced nursing*, 67(2):408–417, 2011.

[87] J. W. Han, Y. D. Ahn, W.-S. Kim, C. M. Shin, S. J. Jeong, Y. S. Song, Y. J. Bae, and J.-M. Kim. Psychiatric manifestation in patients with parkinson's disease. *Journal of Korean medical science*, 33(47), 2018.

[88] G. M. Harari, S. D. Gosling, R. Wang, F. Chen, Z. Chen, and A. T. Campbell. Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. *Computers in Human Behavior*, 67:129–138, 2017.

[89] G. M. Harari, S. R. Müller, C. Stachl, R. Wang, W. Wang, M. Bühner, P. J. Rentfrow, A. T. Campbell, and S. D. Gosling. Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of personality and social psychology*, 2019.

[90] G. M. Harari, S. S. Vaid, S. R. Müller, C. Stachl, Z. Marrero, R. Schoedel, M. Bühner, and S. D. Gosling. Personality sensing for theory development and assessment in the digital age. *European Journal of Personality*, 2020.

[91] R. Harré and P. F. Secord. The explanation of social behaviour. 1972.

[92] A. Hawton, C. Green, A. P. Dickens, S. H. Richards, R. S. Taylor, R. Edwards, C. J. Greaves, and J. L. Campbell. The impact of social isolation on the health status and health-related quality of life of older people. *Quality of Life Research*, 20(1):57–67, 2011.

[93] G. R. Hayes, E. S. Poole, G. Iachello, S. N. Patel, A. Grimes, G. D. Abowd, and K. N. Truong. Physical, social, and experiential knowledge in pervasive computing environments. *IEEE Pervasive Computing*, 6(4):56–63, 2007.

[94] J. V. Hernandez. *Unobtrusive and Personalised Monitoring of Parkinson's Disease Using Smartphones*. The University of Manchester (United Kingdom), 2019.

[95] R. A. Hill and R. I. Dunbar. Social network size in humans. *Human nature*, 14(1):53–72, 2003.

[96] J. Holt-Lunstad, T. B. Smith, M. Baker, T. Harris, and D. Stephenson. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. *Perspectives on psychological science*, 10(2):227–237, 2015.

[97] K. Horvath, Z. Aschermann, M. Kovacs, A. Makkos, M. Harmat, J. Janszky, S. Komoly, K. Karadi, and N. Kovacs. Changes in quality of life in parkinson's disease: how large must they be to be relevant? *Neuroepidemiology*, 48(1-2):1–8, 2017.

[98] K. Huckvale, S. Venkatesh, and H. Christensen. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *NPJ digital medicine*, 2(1):1–11, 2019.

[99] A. Ibrahim, S. Clinch, and S. Harper. Recognising intrinsic motivation using smartphone trajectories. *International Journal of Human-Computer Studies*, 153:102650, 2021.

[100] S. H. Jain, B. W. Powers, J. B. Hawkins, and J. S. Brownstein. The digital phenotype. *Nature biotechnology*, 33(5):462, 2015.

[101] J. Jankovic. Parkinson's disease: clinical features and diagnosis. *Journal of neurology, neurosurgery & psychiatry*, 79(4):368–376, 2008.

[102] C. Jenkinson, R. Fitzpatrick, V. Peto, R. Greenhall, and N. Hyman. The parkinson's disease questionnaire (pdq-39): development and validation of a parkinson's disease summary index score. *Age and ageing*, 26(5):353–357, 1997.

[103] C. Jenkinson, R. Fitzpatrick, V. Peto, R. Greenhall, and N. Hyman. The pdq-8: development and validation of a short-form parkinson's disease questionnaire. *Psychology and Health*, 12(6):805–814, 1997.

[104] G. K. Jonsson. Personality and self-esteem in social interaction. *From communication to presence*, pages 189–205, 2006.

[105] L. V. Kalia and A. E. Lang. Parkinson's disease. *The Lancet*, 386(9996):896–912, 2015.

[106] A. E. Kazdin. Unobtrusive measures in behavioral assessment. *Journal of Applied Behavior Analysis*, 12(4):713–724, 1979.

[107] A. A. Kehagia. Neuropsychiatric symptoms in parkinson's disease: beyond complications. *Frontiers in Psychiatry*, 7:110, 2016.

[108] A. A. Kehagia, R. A. Barker, and T. W. Robbins. Neuropsychological and clinical heterogeneity of cognitive impairment and dementia in patients with parkinson's disease. *The Lancet Neurology*, 9(12):1200–1213, 2010.

[109] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad. Mobile phone sensing systems: A survey. *IEEE Communications Surveys & Tutorials*, 15(1):402–427, 2012.

[110] M. Khwaja, S. S. Vaid, S. Zannone, G. M. Harari, A. A. Faisal, and A. Matic. Modeling personality vs. modeling personalidad: In-the-wild mobile data analysis in five countries suggests cultural impact on personality models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–24, 2019.

[111] C. Killeen. Loneliness: an epidemic in modern society. *Journal of advanced nursing*, 28(4):762–770, 1998.

[112] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila. Towards rich mobile phone datasets: Lausanne data collection campaign. *Proc. ICPS, Berlin*, 68, 2010.

[113] P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower. Exploring privacy concerns about personal sensing. In *International Conference on Pervasive Computing*, pages 176–183. Springer, 2009.

[114] P. Kostopoulos, A. I. Kyritsis, M. Deriaz, and D. Konstantas. Stress detection using smart phone data. pages 340–351, Cham, 2017. Springer International Publishing.

[115] P. Kostopoulos, A. I. Kyritsis, M. Deriaz, and D. Konstantas. Stress detection using smart phone data. In *eHealth 360°*, pages 340–351. Springer, 2017.

[116] A. Kumar and N. Epley. It's surprisingly nice to hear you: Misunderstanding the impact of communication media can lead to suboptimal choices of how to connect with others. *Journal of Experimental Psychology: General*, 2020.

[117] E. Kuosmanen, F. Wolling, J. Vega, V. Kan, Y. Nishiyama, S. Harper, K. Van Laerhoven, S. Hosio, and D. Ferreira. Smartphone-based monitoring of parkinson disease: Quasi-experimental study to quantify hand tremor severity and medication effectiveness. *JMIR mHealth and uHealth*, 8(11):e21543, 2020.

[118] K.-C. Lan and W.-Y. Shih. Early diagnosis of parkinson's disease using a smartphone. *Procedia Computer Science*, 34:305–312, 2014.

[119] N. D. Lane, M. Lin, M. Mohammod, X. Yang, H. Lu, G. Cardone, S. Ali, A. Doryab, E. Berke, A. T. Campbell, et al. Bewell: Sensing sleep, physical activities and social interactions to promote wellbeing. *Mobile Networks and Applications*, 19(3):345–359, 2014.

[120] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. A survey of mobile phone sensing. *IEEE Communications magazine*, 48(9):140–150, 2010.

[121] A. C. Leon, L. L. Davis, and H. C. Kraemer. The role and interpretation of pilot studies in clinical research. *Journal of psychiatric research*, 45(5):626–629, 2011.

[122] T. C. Leonard. *Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness*, volume 19. Yale University Press, 2008.

[123] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma. Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, pages 1–10, 2008.

[124] J. Liddle, D. Ireland, S. J. McBride, S. G. Brauer, L. M. Hall, H. Ding, M. Karunanithi, P. W. Hodges, D. Theodoros, P. A. Silburn, et al. Measuring the lifespace of people with parkinson's disease using smartphones: proof of principle. *JMIR mHealth and uHealth*, 2(1):e2799, 2014.

[125] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong. Can your smartphone infer your mood. In *PhoneSense workshop*, pages 1–5, 2011.

[126] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong. Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pages 389–402, 2013.

[127] M. Linares-Del Rey, L. Vela-Desojo, and R. Cano-de La Cuerda. Mobile phone applications in parkinson's disease: A systematic review. *Neurología (english edition)*, 34(1):38–54, 2019.

[128] F. Lipsmeier, K. I. Taylor, T. Kilchenmann, D. Wolf, A. Scotland, J. Schjodt-Eriksen, W.-Y. Cheng, I. Fernandez-Garcia, J. Siebourg-Polster, L. Jin, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 parkinson's disease clinical trial. *Movement Disorders*, 33(8):1287–1297, 2018.

[129] M. A. Little. Smartphones for remote symptom monitoring of parkinson's disease. *Journal of Parkinson's Disease*, (Preprint):1–5, 2021.

[130] W. Little, R. McGivern, and N. Kerins. *Introduction to Sociology-2nd Canadian Edition*. BC Campus, 2016.

[131] S. Liu, Y. Jiang, and A. Striegel. Face-to-face proximity estimationusing bluetooth on smartphones. *IEEE Transactions on Mobile Computing*, 13(4):811–823, 2013.

[132] S. Liu and A. Striegel. Accurate extraction of face-to-face proximity using smartphones and bluetooth. In *2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*, pages 1–5. IEEE, 2011.

[133] J. L. Locher, C. S. Ritchie, D. L. Roth, P. S. Baker, E. V. Bodner, and R. M. Allman. Social isolation, support, and capital and nutritional risk in an older sample: ethnic and gender differences. *Social science & medicine*, 60(4):747–761, 2005.

[134] H.-I. Ma, M. Saint-Hilaire, C. A. Thomas, and L. Tickle-Degnen. Stigma as a key determinant of health-related quality of life in parkinson's disease. *Quality of Life Research*, 25(12):3037–3045, 2016.

[135] A. Madan, M. Cebrian, D. Lazer, and A. Pentland. Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 291–300, 2010.

[136] A. Madan, K. Farrahi, D. Gatica-Perez, and A. S. Pentland. Pervasive sensing to model political opinions in face-to-face networks. In *International Conference on Pervasive Computing*, pages 214–231. Springer, 2011.

[137] A. Madan, S. T. Moturu, D. Lazer, and A. S. Pentland. Social sensing: obesity, unhealthy eating and exercise in face-to-face networks. In *Wireless Health 2010*, pages 104–110. ACM, 2010.

[138] W. Maetzler, J. Klucken, and M. Horne. A clinical view on the development of technology-based tools in managing parkinson's disease. *Movement Disorders*, 31(9):1263–1271, 2016.

[139] M. Maffoni, A. Giardini, A. Pierobon, D. Ferrazzoli, and G. Frazzitta. Stigma experienced by parkinson's disease patients: a descriptive review of qualitative studies. *Parkinson's Disease*, 2017, 2017.

[140] C. Marsden, J. Parkes, and N. Quinn. Fluctuations of disability in parkinson's disease–clinical aspects. In *Movement disorders*, pages 96–122. Elsevier, 1981.

[141] P. Martinez-Martin, C. Rodriguez-Blazquez, M. M. Kurtis, K. R. Chaudhuri, and N. V. Group. The impact of non-motor symptoms on health-related quality of life of patients with parkinson's disease. *Movement Disorders*, 26(3):399–406, 2011.

[142] M. R. Mehl, J. W. Pennebaker, D. M. Crow, J. Dabbs, and J. H. Price. The electronically activated recorder (ear): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, 33(4):517–523, 2001.

[143] C. Meurisch, T. Hussain, A. Gogel, B. Schmidt, I. Schweizer, and M. Mühlhäuser. A spatiotemporal approach for social situation recognition. In *International Conference on Mobile Computing, Applications, and Services*, pages 309–316. Springer, 2015.

[144] S. Michie, M. M. Van Stralen, and R. West. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*, 6(1):42, 2011.

[145] N. Miller, S. Andrew, E. Noble, and M. Walshe. Changing perceptions of self as a communicator in parkinson's disease: a longitudinal follow-up study. *Disability and rehabilitation*, 33(3):204–210, 2011.

[146] N. Miller, E. Noble, D. Jones, and D. Burn. Life with communication changes in parkinson's disease. *Age and ageing*, 35(3):235–239, 2006.

[147] E. Mioshi, K. Dawson, J. Mitchell, R. Arnold, and J. R. Hodges. The addenbrooke's cognitive examination revised (ace-r): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, 21(11):1078–1085, 2006.

[148] T. M. Mitchell. Mining our reality. *Science*, 326(5960):1644–1645, 2009.

[149] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine*, 151(4):264–269, 2009.

[150] D. C. Mohr, M. Zhang, and S. M. Schueller. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual review of clinical psychology*, 13:23–47, 2017.

[151] Y. Molina, S. W. Choi, D. Cella, and D. Rao. The stigma scale for chronic illnesses 8-item version (ssci-8): development, validation and use across neurological conditions. *International journal of behavioral medicine*, 20(3):450–460, 2013.

[152] B. Mønsted, A. Mollgaard, and J. Mathiesen. Phone-based metric as a predictor for basic personality traits. *Journal of Research in Personality*, 74:16–22, 2018.

[153] C. Montag, K. Błaszkiewicz, B. Lachmann, I. Andone, R. Sariyska, B. Trendafilov, M. Reuter, and A. Markowetz. Correlating personality and actual phone usage. *Journal of Individual Differences*, 2014.

[154] O. Moore, C. Peretz, and N. Giladi. Freezing of gait affects quality of life of peoples with parkinson's disease beyond its relationships with mobility and gait. *Movement disorders: official journal of the Movement Disorder Society*, 22(15):2192–2195, 2007.

[155] S. T. Moturu, I. Khayal, N. Aharony, W. Pan, and A. Pentland. Using social sensing to understand the links between sleep, mood, and sociability. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 208–214. IEEE, 2011.

[156] P. Narme, H. Mouras, M. Roussel, C. Duru, P. Krystkowiak, and O. Godefroy. Emotional and cognitive social processes are impaired in parkinson's disease and are related to behavioral disorders. *Neuropsychology*, 27(2):182, 2013.

[157] A. W. Nguyen, R. J. Taylor, H. O. Taylor, and L. M. Chatters. Objective and subjective social isolation and psychiatric disorders among african americans. *Clinical Social Work Journal*, 48(1):87–98, 2020.

[158] D. H. Nguyen, A. Kobsa, and G. R. Hayes. An empirical investigation of concerns of everyday tracking and recording technologies. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 182–191, 2008.

[159] NICE. *Psychosis and schizophrenia in adults: prevention and management*. NICE, 2014.

[160] G. Nijhof. Parkinson's disease as a problem of shame in public appearance. *Sociology of Health & Illness*, 17(2):193–205, 1995.

[161] Ofcom. Do you personally use a smartphone?. statista - the statistics portal, 2019.

[162] S. Ofcom. Smartphone ownership penetration in the united kingdom (uk) in 2008-2020, by demographics. in statista - the statistics portal, 2020.

[163] R. O. Oloritun, A. Madan, A. Pentland, and I. Khayal. Evolution of social encounters in ad-hoc mobile face-to-face interaction networks. In *2012 International Conference on Social Informatics*, pages 192–198. IEEE, 2012.

[164] J.-P. Onnela and S. L. Rauch. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*, 41(7):1691–1696, 2016.

[165] A. Page, N. Yung, P. Auinger, C. Venuto, A. Glidden, E. Macklin, L. Omberg, M. A. Schwarzschild, and E. R. Dorsey. A smartphone application as an exploratory endpoint in a phase 3 parkinson's disease clinical trial: A pilot study. *Digital Biomarkers*, 6(1):1–8, 2022.

[166] R. Pahwa and K. E. Lyons. *Handbook of Parkinson's disease*. Crc Press, 2013.

[167] N. Palaghias, S. A. Hoseinitabatabaei, M. Nati, A. Gluhak, and K. Moessner. A survey on mobile social signal processing. *ACM Computing Surveys (CSUR)*, 48(4):57, 2016.

[168] M. D. Pell, L. Monetta, K. Rothermich, S. A. Kotz, H. S. Cheang, and S. McDonald. Social perception in adults with parkinson's disease. *Neuropsychology*, 28(6):905, 2014.

[169] A. Perrin. Social media usage. *Pew research center*, pages 52–68, 2015.

[170] V. Peto, C. Jenkinson, and R. Fitzpatrick. Pdq-39: a review of the development, validation and application of a parkinson's disease quality of life questionnaire and its associated measures. *Journal of neurology*, 245(1):S10–S14, 1998.

[171] S. Pinto, C. Ozsancak, E. Tripoliti, S. Thobois, P. Limousin-Dowsey, and P. Auzou. Treatments for dysarthria in parkinson's disease. *The Lancet Neurology*, 3(9):547–556, 2004.

[172] T. K. Pitcairn, S. Clemie, J. M. Gray, and B. Pentland. Non-verbal cues in the self-presentation of parkinsonian patients. *British Journal of Clinical Psychology*, 29(2):177–184, 1990.

[173] G. Pluck and R. Brown. Apathy in parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 73(6):636–642, 2002.

[174] M. Politis, K. Wu, S. Molloy, P. G. Bain, K. R. Chaudhuri, and P. Piccini. Parkinson's disease symptoms: the patient's perspective. *Movement Disorders*, 25(11):1646–1651, 2010.

[175] S. Porcelli, N. Van Der Wee, S. van der Werff, M. Aghajani, J. C. Glennon, S. van Heukelum, F. Mogavero, A. Lobo, F. J. Olivera, E. Lobo, et al. Social brain, social dysfunction and social withdrawal. *Neuroscience & Biobehavioral Reviews*, 97:10–33, 2019.

[176] G. Postolache and O. Postolache. Smartphone sensing technologies for tailored parkinson's disease diagnosis and monitoring. *Mobile solutions and their usefulness in everyday life*, pages 251–273, 2019.

[177] A. Prasad, J. Sorber, T. Stablein, D. Anthony, and D. Kotz. Understanding sharing preferences and behavior for mhealth devices. In *Proceedings of the 2012 ACM workshop on Privacy in the electronic society*, pages 117–128, 2012.

[178] M. Prenger, R. Madray, K. Van Hedger, M. Anello, and P. A. MacDonald. Social symptoms of parkinson's disease. *Parkinson's Disease*, 2020, 2020.

[179] J. Prince, S. Arora, and M. de Vos. Big data in parkinson's disease: using smartphones to remotely detect longitudinal disease phenotypes. *Physiological measurement*, 39(4):044005, 2018.

[180] T. Pringsheim, N. Jette, A. Frolkis, and T. D. Steeves. The prevalence of parkinson's disease: a systematic review and meta-analysis. *Movement disorders*, 29(13):1583–1590, 2014.

[181] J. O. Prochaska, C. A. Redding, K. E. Evers, et al. The transtheoretical model and stages of change. *Health behavior: Theory, research, and practice*, pages 125–148, 2015.

[182] G. Pulekar and E. Agu. Autonomously sensing loneliness and its interactions with personality traits using smartphones. In *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*, pages 134–137. IEEE, 2016.

[183] M. Rabbi, A. Pfammatter, M. Zhang, B. Spring, and T. Choudhury. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. *JMIR mHealth and uHealth*, 3(2):e42, 2015.

[184] C. Ramaker, J. Marinus, A. M. Stiggelbout, and B. J. Van Hilten. Systematic evaluation of rating scales for impairment and disability in parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society*, 17(5):867–876, 2002.

[185] S. Rigby, E. Thornton, S. Tedman, F. Burchardt, C. Young, and C. Dougan. Quality of life assessment in mnd: development of a social withdrawal scale. *Journal of the neurological sciences*, 169(1-2):26–34, 1999.

[186] S. Riggare and M. Hägglund. Precision medicine in parkinson's disease–exploring patient-initiated self-tracking. *Journal of Parkinson's Disease*, 8(3):441–446, 2018.

[187] R. J. Rummel. Social behavior and interaction. *Rummel, RJ, Conflict and War*, 2, 1976.

[188] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, D. C. Mohr, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet research*, 17(7):e4273, 2015.

[189] C. Saldert and M. Bauer. Multifaceted communication problems in everyday conversations involving people with parkinson's disease. *Brain sciences*, 7(10):123, 2017.

[190] A. Samii, J. G. Nutt, and B. R. Ransom. Parkinson's disease. *The Lancet*, 363(9423):1783–1793, 2004.

[191] S. Sapir. Multiple factors are involved in the dysarthria associated with parkinson's disease: a review with implications for clinical practice and research. *Journal of Speech, Language, and Hearing Research*, 57(4):1330–1343, 2014.

[192] M. Schmid Mast, D. Gatica-Perez, D. Frauendorfer, L. Nguyen, and T. Choudhury. Social sensing for psychology: Automated interpersonal behavior assessment. *Current Directions in Psychological Science*, 24(2):154–160, 2015.

[193] R. Schoedel, F. Pargent, Q. Au, S. T. Völkel, T. Schuwerk, M. Bühner, and C. Stachl. To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day–night behaviour patterns. *European Journal of Personality*, 2020.

[194] A. Schrag, M. Jahanshahi, and N. Quinn. How does parkinson's disease affect quality of life? a comparison with quality of life in the general population. *Movement disorders: official journal of the Movement Disorder Society*, 15(6):1112–1118, 2000.

[195] T. Schuwerk, L. J. Kaltefleiter, J.-Q. Au, A. Hoesl, and C. Stachl. Enter the wild: Autistic traits and their relationship to mentalizing and social interaction in everyday life. *Journal of autism and developmental disorders*, 49(10):4193–4208, 2019.

[196] I. Schweizer, R. Bärtl, B. Schmidt, F. Kaup, and M. Mühlhäuser. Kraken. me mobile: the energy footprint of mobile tracking. In *6th International Conference on Mobile Computing, Applications and Services*, pages 82–89. IEEE, 2014.

[197] P. Sedgwick and N. Greenwood. Understanding the hawthorne effect. *Bmj*, 351, 2015.

[198] T. E. Seeman, T. M. Lusignolo, M. Albert, and L. Berkman. Social relationships, social support, and patterns of cognitive aging in healthy, high-functioning older adults: Macarthur studies of successful aging. *Health psychology*, 20(4):243, 2001.

[199] S. Servia-Rodríguez, K. K. Rachuri, C. Mascolo, P. J. Rentfrow, N. Lathia, and G. M. Sandstrom. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the 26th International Conference on World Wide Web*, pages 103–112, 2017.

[200] L. M. Shulman. Understanding disability in parkinson's disease. *Movement disorders*, 25(S1):S131–S135, 2010.

[201] A. M. Simons, T. Beltramo, G. Blalock, and D. I. Levine. Using unobtrusive sensors to measure and minimize hawthorne effects: Evidence from cookstoves. *Journal of Environmental Economics and Management*, 86:68–80, 2017.

[202] V. K. Singh and R. R. Agarwal. Cooperative phoneotypes: exploring phone-based behavioral markers of cooperation. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 646–657. ACM, 2016.

[203] M. A. Soleimani, R. Negarandeh, F. Bastani, and R. Greysen. Disrupted social connectedness in people with parkinson's disease. *British Journal of Community Nursing*, 19(3):136–141, 2014.

[204] A. Solomon, A. Bar, C. Yanai, B. Shapira, and L. Rokach. Predict demographic information using word2vec on spatial trajectories. In *Proceedings of the 26th conference on user Modeling, adaptation and personalization*, pages 331–339, 2018.

[205] H. Son, W. S. Park, and H. Kim. Mobility monitoring using smart technologies for parkinson's disease in free-living environment. *Collegian*, 25(5):549–560, 2018.

[206] A. Soundy, B. Stubbs, and C. Roskell. The experience of parkinson's disease: a systematic review and meta-ethnography. *The Scientific World Journal*, 2014, 2014.

[207] D. Spruijt-Metz and W. Nilsen. Dynamic models of behavior for just-in-time adaptive interventions. *IEEE Pervasive Computing*, 13(3):13–17, 2014.

[208] C. Stachl, Q. Au, R. Schoedel, D. Buschek, S. Völkel, T. Schuwerk, M. Oldemeier, T. Ullmann, H. Hussmann, B. Bischl, et al. Behavioral patterns in smartphone usage predict big five personality traits. 2019.

[209] C. Stachl, Q. Au, R. Schoedel, S. D. Gosling, G. M. Harari, D. Buschek, S. T. Völkel, T. Schuwerk, M. Oldemeier, T. Ullmann, et al. Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, 117(30):17680–17687, 2020.

[210] C. Stamate, G. D. Magoulas, S. Küppers, E. Nomikou, I. Daskalopoulos, A. Jha, J. Pons, J. Rothwell, M. U. Luchini, T. Moussouri, et al. The cloudupdrs app: A medical device for the clinical assessment of parkinson's disease. *Pervasive and mobile computing*, 43:146–166, 2018.

[211] S. E. Starkstein, H. S. Mayberg, T. Preziosi, P. Andrezejewski, R. Leiguarda, R. Robinson, et al. Reliability, validity, and clinical correlates of apathy in parkinson's disease. *J Neuropsychiatry Clin Neurosci*, 4(2):134–139, 1992.

[212] S. E. Starkstein, T. J. Preziosi, M. L. Berthier, P. L. Bolduc, H. S. Mayberg, and R. G. Robinson. Depression and cognitive impairment in parkinson's disease. *Brain*, 112(5):1141–1153, 1989.

[213] S. E. Starkstein, T. J. Preziosi, A. W. Forrester, and R. G. Robinson. Specificity of affective and autonomic symptoms of depression in parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 53(10):869–873, 1990.

[214] Statista. Which of the following electronic devices does your household own?, 2020.

[215] Statista. Which of the following electronic devices does your household own? *www.statista.com/forecasts/997925/consumer-electronics-ownership-in-the-uk*, 2020.

[216] A. Steptoe, A. Shankar, P. Demakakos, and J. Wardle. Social isolation, loneliness, and all-cause mortality in older men and women. *Proceedings of the National Academy of Sciences*, 110(15):5797–5801, 2013.

[217] E. W. Steyerberg, M. J. Eijkemans, and J. D. F. Habbema. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *Journal of clinical epidemiology*, 52(10):935–942, 1999.

[218] F. Stocchi, P. Martínez-Martin, and H. Reichmann. Quality of life in parkinson's disease-patient, clinical and research perspectives. *European Neurological Review*, 9(1):12–18, 2014.

[219] M. Strugar. 30+ smartphone usage statistics for the uk [2021], 2021.

[220] A. Sutcliffe, R. Dunbar, J. Binder, and H. Arrow. Relationships and the social brain: integrating psychological and evolutionary perspectives. *British journal of psychology*, 103(2):149–168, 2012.

[221] S. Sveinbjornsdottir. The clinical symptoms of parkinson's disease. *Journal of neurochemistry*, 139:318–324, 2016.

[222] K. Takahashi, N. Kamide, M. Suzuki, and M. Fukuda. Quality of life in people with parkinson's disease: the relevance of social relationships and communication. *Journal of Physical Therapy Science*, 28(2):541–546, 2016.

[223] T. B. I. Team. Covid-19, Apr 2020.

[224] T. B. I. Team. Using behavioural insights to create a covid-19 text service for the nhs, 2020.

[225] A. R. Teo, H. Choi, and M. Valenstein. Social relationships and depression: ten-year follow-up from a nationally representative study. *PloS one*, 8(4):e62396, 2013.

[226] J. Torous, M. V. Kiang, J. Lorme, and J.-P. Onnela. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR mental health*, 3(2):e16, 2016.

[227] J. Torous, P. Staples, I. Barnett, L. R. Sandoval, M. Keshavan, and J.-P. Onnela. Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with schizophrenia. *NPJ digital medicine*, 1(1):1–9, 2018.

[228] F. Tsapeli and M. Musolesi. Investigating causality in human behavior from smartphone sensor data: a quasi-experimental approach. *EPJ Data Science*, 4(1):24, 2015.

[229] X.-J. Tu, W.-J. Hwang, S.-P. Hsu, and H.-I. Ma. Responsiveness of the short-form health survey and the parkinson's disease questionnaire in patients with parkinson's disease. *Health and quality of life outcomes*, 15(1):1–7, 2017.

[230] P. UK. the parkinson's disease questionnaire (pdq-39), 2022.

[231] K. Usher, N. Bhullar, and D. Jackson. Life in the pandemic: Social isolation and mental health. *Journal of Clinical Nursing*, 2020.

[232] J. J. Van Bavel, K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M. J. Crockett, A. J. Crum, K. M. Douglas, J. N. Druckman, et al. Using social

and behavioural science to support covid-19 pandemic response. *Nature Human Behaviour*, pages 1–12, 2020.

[233] N. J. Van der Wee, A. C. Bilderbeck, M. Cabello, J. L. Ayuso-Mateos, I. M. Saris, E. J. Giltay, B. W. Penninx, C. Arango, A. Post, and S. Porcelli. Working definitions, subjective and objective assessments and experimental paradigms in a study exploring social withdrawal in schizophrenia and alzheimer's disease. *Neuroscience & Biobehavioral Reviews*, 97:38–46, 2019.

[234] J. Vega, S. Couth, E. Poliakoff, S. Kotz, M. Sullivan, C. Jay, M. Vigo, and S. Harper. Back to analogue: Self-reporting for parkinson's disease. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13, 2018.

[235] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009.

[236] F. Wahle, T. Kowatsch, E. Fleisch, M. Rufer, and S. Weidt. Mobile sensing and support for people with depression: a pilot trial in the wild. *JMIR mHealth and uHealth*, 4(3):e111, 2016.

[237] H. Wang and S. Marsella. Assessing personality through objective behavioral sensing. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 131–137. IEEE, 2017.

[238] R. Wang, M. S. Aung, S. Abdullah, R. Brian, A. T. Campbell, T. Choudhury, M. Hauser, J. Kane, M. Merrill, E. A. Scherer, et al. Crosscheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 886–897, 2016.

[239] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pages 3–14. ACM, 2014.

[240] R. Wang, W. Wang, M. S. Aung, D. Ben-Zeev, R. Brian, A. T. Campbell, T. Choudhury, M. Hauser, J. Kane, E. A. Scherer, et al. Predicting symptom

trajectories of schizophrenia using mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–24, 2017.

[241] W. Wang, G. M. Harari, R. Wang, S. R. Müller, S. Mirjafari, K. Masaba, and A. T. Campbell. Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–21, 2018.

[242] B. Wee, A. Sutherland, and J. Stickland. Can video consultations replace face to face interviews? *BMJ Supportive and Palliative Care*.

[243] G. C. Wenger, R. Davies, S. Shahtahmasebi, and A. Scott. Social isolation and loneliness in old age: review and model refinement. *Ageing & Society*, 16(3):333–358, 1996.

[244] Wikipedia. Coronavirus disease 2019 — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Coronavirus%20disease%202019&oldid=956033455`, 2020. [Online; accessed 11-May-2020].

[245] Wikipedia. Mobile phone based sensing software, 2021.

[246] Y. Winter, S. von Campenhausen, J. Gasser, K. Seppi, J.-P. Reese, K.-P. Pfeiffer, K. Bötzel, W. H. Oertel, R. Dodel, and W. Poewe. Social and clinical determinants of quality of life in parkinson's disease in austria: a cohort study. *Journal of neurology*, 257(4):638–645, 2010.

[247] C. Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, page 38. Citeseer, 2014.

[248] A. Wootton, N. J. Starkey, and C. C. Barber. Unmoving and unmoved: experiences and consequences of impaired non-verbal expressivity in parkinson's patients and their spouses. *Disability and Rehabilitation*, 41(21):2516–2527, 2019.

[249] P. Xu, M. Dredze, and D. A. Broniatowski. The twitter social mobility index: Measuring social distancing practices from geolocated tweets. *arXiv preprint arXiv:2004.02397*, 2020.

[250] Z. Yan, J. Yang, and E. M. Tapia. Smartphone bluetooth based social sensing. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 95–98, 2013.

[251] Y. C. Yang, C. Boen, K. Gerken, T. Li, K. Schorpp, and K. M. Harris. Social relationships and physiological determinants of longevity across the human life span. *Proceedings of the National Academy of Sciences*, 113(3):578–583, 2016.

[252] J. A. Yesavage. Geriatric depression scale. *Psychopharmacol Bull*, 24(4):709–711, 1988.

[253] Z. Yu, X. Zhou, D. Zhang, G. Schiele, and C. Becker. Understanding social relationship evolution by using real-world sensing data. *World Wide Web*, 16(5-6):749–762, 2013.

[254] A. Zhan, M. A. Little, D. A. Harris, S. O. Abiola, E. Dorsey, S. Saria, and A. Terzis. High frequency remote monitoring of parkinson's disease via smartphone: Platform overview and medication response detection. *arXiv preprint arXiv:1601.00960*, 2016.

[255] A. Zhan, S. Mohan, C. Tarolli, R. B. Schneider, J. L. Adams, S. Sharma, M. J. Elson, K. L. Spear, A. M. Glidden, M. A. Little, et al. Using smartphones and machine learning to quantify parkinson disease severity: the mobile parkinson disease score. *JAMA neurology*, 75(7):876–880, 2018.

[256] H. Zhang, G. Guo, C. Song, C. Xu, K. Cheung, J. Alexis, H. Li, D. Li, K. Wang, and W. Xu. Pdlens: smartphone knows drug effectiveness among parkinson's via daily-life activity fusion. In *Proceedings of the 26th annual international conference on mobile computing and networking*, pages 1–14, 2020.

[257] H. Zhang, A. Ibrahim, B. Parsia, E. Poliakoff, and S. Harper. Passive social sensing with smartphone: A systematic review. submitted, 2020.

[258] H. Zhang, B. Parsia, E. Poliakoff, and S. Harper. Exploring social withdrawal with smartphones in people with parkinson's disease: A longitudinal study. submitted, 2020.

[259] H. Zhang, B. Parsia, E. Poliakoff, and S. Harper. Monitoring social withdrawal with smartphones in people with parkinson's disease and the impact of covid19. submitted, 2021.

[260] H. Zhang, C. Xu, H. Li, A. S. Rathore, C. Song, Z. Yan, D. Li, F. Lin, K. Wang, and W. Xu. Pdmove: Towards passive medication adherence monitoring of parkinson's disease using smartphone-based gait assessment. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 3(3):1–23, 2019.

Appendix A

Study Protocol

A.1 Recruitment

After the ethics approval, the participant recruitment campaign started. Our advertisements were distributed through the Parkinson's research community and the Parkinson's UK website. Potential candidates were also contacted via email and phone. Participants are required to be:

- Clinically diagnosed with idiopathic Parkinson's.
- Fluent English speakers aged between 45 - 80.
- Using Android smartphones
- Not suffering or have suffered from any major physical or neurological conditions

A.2 Ground truth

We are a longitudinal study and will collect a large amount of social behaviour data from the smartphone, and we want to correlate the data from the smartphone and the ground truth of existing technologies – questionnaires. Of course, we want the ground truth as frequent as we can to calibrate the data we collected, but it will put much burden on patients. Besides, we don't have enough time to take these scales for each patient at a high frequency.

The Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [76] is the gold standard in clinical assessment to quantify PD's overall progression. It is the most commonly used scale in the clinical study [184]. By interview and clinical observations, MDS-UPDRS can comprehensively track the longitudinal course of PD in six different parts, including non-motor experiences of daily living (13 items), motor experiences of daily living (13 items), motor symptoms (33 items), and motor complications (6 items). Each item is ranked on a scale from 0 to 4 (Normal, Slight, Mild, Moderate, Severe).

Parkinson's disease questionnaire (PDQ-39) citejenkinson1997parkinson is a Parkinson's disease-specific quality of life questionnaire. It has been developed and thoroughly tested for reliability and validity. It is a reliable, valid, responsive, acceptable, feasible tool and widely used [229]. It has eight subscales designed to probe levels of mobility (10 items), activities of daily living (six items), emotional well-being (six items), stigma (four items), social support (three items), cognition (four items), communication (three items) and bodily discomfort (three items). Participants are asked to tick the appropriate response from 'never', 'occasionally', 'sometimes', 'often' and 'always/ cannot do'.

The Addenbrooke's cognitive examination revised (ACE-R) [147] assesses cognitive skills in five subdomains: orientation/attention, memory, verbal fluency, language, and visuospatial. It is applied at the beginning of the study to screen participants to check if they have cognitive impairment. It will also be applied in every home visit to assess the cognitive decline of participants.

The 8-item Stigma Scale for Chronic Illness (SSCI) [151] was developed to measure the stigma experienced by people with chronic neurological disorders, including PD [Rao et al., 2009]. It contains two subscales: felt stigma and enacted stigma. Each item is rated as 1 = never, 2 = rarely, 3 = sometimes, 4 = often, and 5 = always. A higher score indicates a higher frequency of experiencing stigma.

The 15-item Geriatric Depression Scale (GDS) [252] is a self-report measure to detect depressive symptoms in older adults and has been wildly used in people with PD. The GDS is reported to have adequate discriminant validity for a diagnosis of depressive disorder at a cutoff of 5, with a higher score meaning more depressive.

The apathy Scale (AS) [211] is the only apathy scale that meets the criteria to be "recommended" by the Movement Disorder Society to measure the apathy status of PD patients. The AS has 14 items, and each question is read by the examiner and has four options of response: "not at all", "slightly", "some", or "a lot". Scores range from 0 to 42; higher scores indicate more severe apathy.

The interpersonal reactivity index (IRI) [46] measures the empathy status of participants. There are 28 items and four subscales in it, perspective-taking (PT), fantasy (FS), empathic concern (EC) and personal distress (PD). PT and EC discuss more feelings about other people, which can reflect the empathy ability for social contact. So only PT and EC subscales of IRI are selected, which only have 14 items. The rating is from Does not describe me well (scored 1) to Describes me very well (scored 5).

Since our study focuses on social withdrawal, a scale for that will be an effective supplement for the ground truth. We could not find one specifically for PD, so we chose Social Withdrawal Scale in motor neurone disease (MND) [185], and it was modified for PD patients. Motor neurone disease has similar symptoms as PD, so it is feasible for modification. This scale is designed to assess social withdrawal from the perspective of the MND patient across four domains of Community, Family, Emotional and Physical Withdrawal. The original scale consists of 24 items scored along a four-point Likert-type response ranging from Does not describe me well (scored 1) to Describes me very well (scored 6).

Julio originally created this diary to trace day-to-day fluctuations of personal PD symptoms [234]. Although four other prototypes were tested, including Bluetooth, physical buttons, NFC and microcontrollers, the paper diary achieved the highest acceptance and compliance. In the original design, participants choose three symptoms that most impact their lives and record them daily. To obtain more granular ratings, a shorter version of the Parkinson's disease questionnaire, PDQ-8 [103], was also added to the diary every week. We also designed a weekly questionnaire to measure participants' social interaction levels. This questionnaire asks participants to rate their social interaction levels from 0 to 10, and the levels include different types of contact: family, friends, acquaintances and strangers. An overall rating is also requested. We followed the implications in the original design for all added items to maintain consistency. All ratings were accomplished by filling in the tiny dot of the corresponding number. An example of the answered weekly diary is shown in Appendix D.

A.3 Assessment visits

All the recruitment took place on an enrol-and-go basis. Once the participant signs the agreement to join the study, we will arrange the first home visit with them. At the beginning of the home visit, we will have a cognitive test with candidates using ACE-R. Patients who score more than 88/100 are regarded as having no cognitive impairment and qualified to participate. Then the monitoring application was installed on their smartphone. Then the battery of scales begins. The home visit could last 2.5 hours. In total, there are six visits for each participant. Two participants missed the last round of scales. In each home visit, used diaries were collected, and new diaries were given to participants. So typically, the length of the diary is two months. During the COVID-19 pandemic, all home visits were suspended. The battery of tests is conducted online alternatively. Diaries were collected and sent by post.

Appendix B

Modified Social Withdrawal Scale

Social Withdrawal Scale

Participant ID: Date:

The following statements inquire about your thoughts and feelings in a variety of situations. For each item, indicate how well it describes you by ticking the appropriate option.

Please read each item carefully before responding and answer as honestly as you can.

	Does not describe me well					**Describes me very well**
	1	**2**	**3**	**4**	**5**	**6**
My social life has been completely unaffected by my condition.						
My condition has brought me closer to my family.						
My participation and involvement with local organisations/societies/clubs, has noticeably decreased since my diagnosis.						
My physical condition prevents me doing all the things that I would like to do.						
I feel able to talk to my family/close friends about my feelings.						
I enjoy the company of my close friends.						
I have continued to use public facilities, (i.e., library, leisure centre etc.) to the same extent as I did prior to Parkinson’s.						
I find it difficult to use the toilet in a public place.						
I find getting washed, dressed and ready to go out very difficult and time consuming.						
I still participate in everything-to the same extend as I did, prior to Parkinson’s.						
My relationships with the significant people in my life have become more strained.						
I spend more time alone than I used to.						
I would like to be surrounded, only by the people who know me and understand about my condition.						
I feel under pressure when I am surrounded by other people.						
I want to go out and do things, as much as before I had Parkinson’s.						

I find it difficult to go out, as getting into a car or onto public transport, poses a real problem for me.						
I no longer use the telephone as much as I used to, prior to Parkinson's.						
I worry about the way other people will react to me.						
I am concerned that others may believe that I am drunk or childlike.						
I am totally reliant upon other people, if I want to go out or do something.						
I spend far more time at home than I used to, prior to Parkinson's.						
I feel embarrassed in public places.						
I am only able to go to places which have adequate wheel chair access.						
I feel confident amongst other people.						

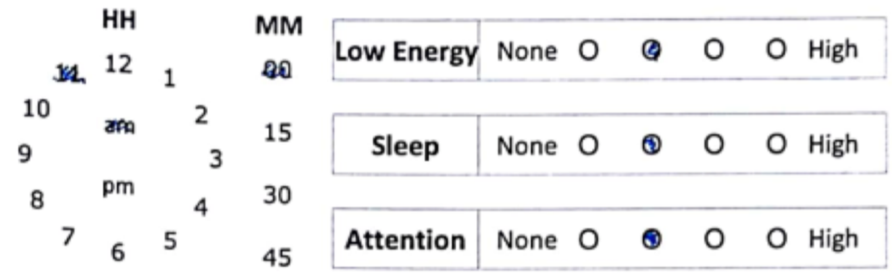
Appendix C

Diary Guidance

Guidance for diary

Aim of the diary:
Capturing **daily symptoms**, **weekly life situation** and **sociality**.
Daily symptoms:
- Each page represents a day for daily symptoms with specific dates.
- Each day has three same sections, bottom two are options.
- Please fill **at least one section per day** at any time.
- If you feel your symptoms fluctuate or just want to report more, please feel free to do so.
- Feel free to write anything such as symbols, text in the notes area of the diary.
- When filling each section, please **fill the circle completely** and **give a strong cross** on **the number of hours, am or pm and minutes**. An example is shown on the right.
- Please make sure you fill each item of the section including the extent of three symptoms and the time you fill it.

Weekly questionnaires:
- After every seven pages, there will be one page asking some questions about your life and another page asking how sociable you have been during the last week.
- Please fill them **on the last day of each week**, specific dates are also provided on the top left.
- When filling questions about your life, please fill the circle thoroughly.
- When filling sociable questions, please fill the circle under the number from 0 to 10 in the box.



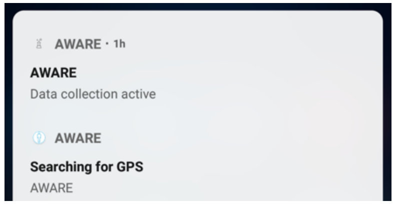
Definitions of social groups:
- **Family and Close Friends**: closest intimates, typically immediate family members and best friends, who provide emotional and behavioral (e.g. financial) support, “people you would seek advice, support or help from in times of severe emotional or financial crisis”).
- **Friends**: reliable friends, whom one can depend on in reciprocal relationships (e.g. friendship in the social sense, protection against harassment, social alliances, distributed childcare), “people whose death would leave your personally devastated”.
- **Acquaintance**: remaining active social relationships, all individual ties with genuine personal relationships (e.g. health professionals, consultant)
- **Strangers:** people you don’t know or are not familiar, (e.g. unknown people providing services such as cashiers of shops, waiters, bartenders)

What if I miss the day?
You can always recall that day or the week and fill it in, just use the time when you fill it. But please add a star (*) next to the date, so we can know that.

Guidance for mobile phone

Basic Information:

- This software is called AWARE and installed as a standalone application with background and accessibility service. Its icon is shown on the right.
- It will always present on the notification panel and is shown below.
- Sometimes, the notification panel will show it is searching for GPS or recording your voice, don't worry, that is AWARE is collecting data.



- **All sensitive data will be processed locally, we will not know the exact name of your contacts, your original sound or any other private information.**

Keep an eye on:

- The software may reduce the performance of your phone slightly and make your battery drain faster.
- Please keep an eye on your battery level and **don't forget to charge your phone** when the battery level is low.
- If it is possible, please connect to Wi-Fi while charging. So the data on your phone can be transferred to us.
- If you happen to see "data collection active" message of **AWARE is not shown in the notification panel**, please **touch the AWARE icon to run it again**.

Settings:

(I will set up these on the first visit, if they fail and you how to fix, please fell free to do so).

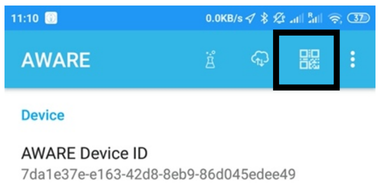
- Please keep Wi-Fi, Bluetooth, GPS and any other sensors on your phone on in the settings all the time.
- Please do not limit the background service of the AWARE.
- Please set the battery saver for AWARE to no restriction.

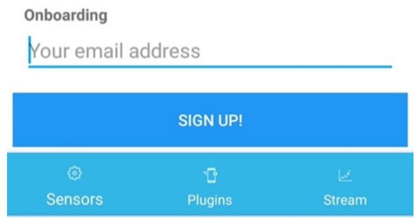
If you want to change to another phone during the study, please let us know so we can arrange a visit to install the AWARE on your new phone and delete it on the old phone.

Rejoin the study:

New technologies sometimes fail. If you are asked to rejoin the study, please follow these steps:

1. Click the AWARE icon from home screen
2. Click the scan icon on the right top corner





3. Scan this QR code above
4. AWARE will ask you to type onboarding email if it scans successfully, please just type the name I sent and click sign up!
5. All done, please allow any rights if it asks you

If you can not finish these step, that's fine. We will sort it out in the next visit, or you can contact me by phone or email.

Any other questions, please contact Heng Zhang ([e-mail], [phone number])

Appendix D

An Example of the Answered Weekly Diary

LOCKDOWN STILL IN OPERATION.

Monday, 25 Jan 2021

How sociable were you during the last week?

Please indicate your extent of social contact of following groups (full definitions of each group can be found in the guidance) <u>during the last week</u>, **0 means never contact, 10 means very high level of contact**.

*Please **fill one circle** for each group*

	0	1	2	3	4	5	6	7	8	9	10
Family and Close Friends (close intimates, typically immediate family members and best friends)	○	○	○	●	○	○	○	○	○	○	○
Friends (reliable friends in reciprocal relationships)	○	○	●	○	○	○	○	○	○	○	○
Acquaintance (all remaining individual ties with genuine relationships, e.g. health professions)	○	○	●	○	○	○	○	○	○	○	○
Strangers (people you don't know, e.g. cashiers of shops, waiters)	○	○	●	○	○	○	○	○	○	○	○
Overall (extent of all your social interactions including strangers)	○	○	●	○	○	○	○	○	○	○	○

Please check that you have filled one circle for each group

Monday, 25 Jan 2021

How was your life during the last week?

Due to having Parkinson's disease
How often <u>during the last week</u> have you

*Please **fill one circle** for each question*

	Never	**Occasionally**	**Sometimes**	**Often**	**Always** or can not do at all
Had difficulty getting around in public?	○	○	●	○	○
Had difficulty dressing yourself?	○	○	●	○	○
Felt depressed?	○	○	●	○	○
Had problems with your close personal relationships?	○	○	●	○	○
Had problems with your concentration, e.g. When reading or watching TV?	○	○	○	●	○
Felt unable to communicate with people properly?	○	○	●	○	○
Had painful muscle cramps or spasms?	○	○	○	○	●
Felt embarrassed in public due to having Parkinson's disease?	○	○	●	○	○

Please check that you have filled on circle for each question