General Discussion

Methodological and statistical perspectives on smartphone-based digital phenotyping
A recent and increasing methodological trend in behavioural phenotyping is the so-called digital phenotyping and is characterised by the longitudinal and digital monitoring of behaviour via smartphones. Due to this increasing trend, several digital phenotyping applications such as the Beiwe App, MONARCA system, Purple Robot, and our own BEHAPP application have been developed for smartphones over the past years. These smartphone applications all have the ability to collect behavioural data in a passive manner by utilizing the user-logs and embedded sensors in smartphones (e.g., global Positioning System (GPS), accelerometer, Wi-Fi, and/or microphone). The data collected by these sensors generate a high-resolution trace of behavioural data which subsequently is used to derive behavioural phenotypes, or so-called digital phenotypes. Currently, this digital phenotyping approach is predominantly employed in the context of psychiatric research with the aim of obtaining a better understanding of the impairments that characterise these disorders in day-to-day life. Subsequently, these behavioural insights in day-to-day life of patients with a psychiatric disorder are used to identify the underlying biological mechanisms involved with these impairments. In addition to psychiatric research, this digital approach can also be employed in other research disciplines to quantify behaviour and potentially provide meaningful insights. With this in mind, we expect that this upcoming research tool will find its way to other research disciplines in the near future.

The rationale behind this novel approach is that the data is collected in a more objective manner relative to traditional assessments of behaviour such as questionnaires, self- or proxy-rated measures and other subjective assessment methods. While it is important to acknowledge that these traditional methods led to numerous important insights, these methods also have their limitations that impede their objectivity and construct validity. The most notable factors that impede the validity of these methods are their dependency on the respondent’s subjective (or the respondent’s proxy) account of behaviour which are invariably obtained post-hoc (i.e. questionnaire measures of behaviour are virtually never real-time). This is particularly relevant for psychiatry, which almost exclusively relies on self-reports of mental health symptoms and behaviour for the diagnosis and treatment of mental disorders. In chapter 2 for example, it was shown that self-reported WHODAS scores in schizophrenia and Alzheimer’s disease patients deviated from proxy reported WHODAS scores by researchers and caregivers, and that these self- versus proxy-related scale score differences were strongly related to disease severity. These results suggest that these biased and subjective reports of symptoms and behaviour are limited to the subjective experience of the patient in a clinical setting and does not reflect behaviour in a natural setting. The smartphone in contrast provides a more objective and ecological trace of behav-
Phenotyping data that is collected in real-time, in the subject’s natural environment, and without the need for any self- or proxy reporting. These features together address some of the most important limitations of traditional assessments of behaviour used for human phenotyping. Also adding to the appeal of digital phenotyping is the relatively low-cost of this approach in terms of scaling, subjects that consent to participation simply have to install and activate an application and the data is collected without any further active input required. This aspect of scalability becomes particularly attractive when we take into account that the majority of people nowadays own a smartphone with a wide range of sensors able to collect relevant behavioural data.

The promise of this approach is that the derived phenotypes may provide unprecedented and unique insights into human behaviour in situ. It is argued that this more objective assessment has the potential to enhance our understanding of the variations in behaviour within specific populations and the underlying biological mechanisms related to these variations. A recent review article identified over 80 peer-reviewed publications since 2015 that used the concept of digital phenotyping to study psychiatric disorders. The authors of this article describe a variety of different studies with promising results in terms of clinical relevance and as a research tool. For example, researchers have demonstrated that phone usage and activity data (i.e. accelerometer and GPS) can be used to identify individuals at risk for depression and anxiety disorders. Additionally, several studies have confirmed that this novel approach has the potential to detect psychiatric relapse. In bipolar disorder for example, changes in activity patterns, smartphone usage, smartphone-based social behaviour, and voice have been used to identify and predict both manic and depressive episodes. It is argued that the self-monitoring of these changes in behaviour could act as an early warning system to prevent relapse and hospitalisation. Also the first results of studies in schizophrenia suggest that the monitoring of behaviour via smartphones can be used for the prediction of relapses and hospitalisation. Given the chronic and recurrent course of most psychiatric disorders, it is likely that with the ecological and objective nature of this smartphone-based approach researchers have the ability to gain a better understanding of these disorders in day-to-day life. Therefore, and in addition to the aforementioned benefits, this approach also has the ability to contribute to the discovery of new treatments for psychiatric patients and evaluate the efficiency of current interventions.

The potential of digital phenotyping in psychiatry and other disciplines is matched by a variety of different challenges that are inherent to this novel approach. Some of these challenges are addressed by preceding literature and mainly focused on the steps needed to apply this approach in clinical practice and their regular-
Methodological and statistical perspectives on smartphone-based digital phenotyping

Here, we will describe several methodological and statistical challenges that need to be addressed in order to employ this approach in research or clinical practice. The challenges described here are based on our experience with our digital phenotyping platform called BEHAPP and until now are often overlooked by preceding literature. The aim of this chapter is to discuss and emphasize the current methodological and statistical limitations of this approach and their potential solutions. In this chapter we (1) discuss the lack of methodological consensus in deriving behavioural phenotypes from this multimodal and complex smartphone sensors data, (2) describe the steps needed to validate smartphone-based monitoring of symptoms as a predictor of relapse in psychiatric disorders, (3) describe recruitment biases related to the severity of symptoms in psychiatric patients, and (4) describe the limitations of the use of location data (i.e. GPS data) in research and introduce a potential alternative.

**DERIVING DIGITAL PHENOTYPES FROM SMARTPHONE DATA**

Deriving these so-called digital phenotypes to study behaviour is achieved by using the behavioural and longitudinal data that is passively collected by smartphone sensors and phone usage logs. This data contains a variety of different modalities, including, but not limited to, location data (i.e. GPS), background noise (i.e. microphone), physical movement data (i.e. accelerometer), Wi-Fi access point data and social behavioural data (text and call logs and social media usage). The premise of digital phenotyping is that this high-resolution and complex trace of behavioural data can be used to generate behavioural phenotypes that are relevant for several research disciplines and potentially, in the future, for clinical practice. For example, in psychiatry this source of data could be used to study the mobility patterns and the degree of social functioning in subjects diagnosed with a psychiatric disorder.

However, a major challenge here is how to generate these relevant behavioural phenotypes by using pre-processing procedures that are validated in the context of smartphone data. This need for validated pre-processing procedures is emphasized by the lack of methodological consensus between studies that utilised smartphone data to extract behavioural phenotypes. Currently, different studies employ a variety of different procedures to generate identical phenotypes in terms of their behavioural interpretation. As a consequence of this variety in methodological procedures the comparison of the results reported by these studies is limited. For example, we identified three experiments\(^8,9,24\) that aimed to study the depressive symptoms in bipolar disorder by using smartphone-based location data. All three experiments utilised a different methodological procedure to derive the same behavioural phenotype (i.e. number of location visited and/or time spent...
at home). Despite that identical phenotypes are derived in terms of their behav-
ioural interpretation, the difference in methodological procedures limits our ability
to compare the results of these three studies.

Standardised frameworks are needed to pre-process the data and derive rel-
evant behavioural phenotypes in order to compare the results of different studies. Recently, we developed a standardised framework for pre-processing and deriv-
ing phenotypes from smartphone-based location data\textsuperscript{25} (chapter 3). In contrast to
preceding literature, this pre-processing procedure is based on methods that are
validated in the context of location data. This framework first generates context-
enriched location data by identifying the stationary (non-moving) and non-station-
ary (movement) states and subsequently clusters those stationary locations that
are recurrent over time (Figure 1). Subsequently, this context-enriched location
data can be used to generate behavioural phenotypes related to several aspects
of mobility. We argue that methodological consensus among researchers that are
applying this approach is essential in order assess the clinical relevance of this
digital approach. To assess this, it is important that the results of various studies
can be compared and reproduced by using comparable methods. The aforemen-
tioned pre-processing framework is solely using the location data collected by
smartphones, additional frameworks are needed to derive phenotypes from other
smartphone modalities such as accelerometer and phone usage data.

A limitation of this approach is that the derived digital phenotypes are currently
generated manually from raw smartphone data. Given the endless number of pos-
sibilities, this process is time-consuming and often driven by domain knowledge
and a-priori expectations about behaviour in specific populations. In order to take
full benefit of the more objective assessment of behaviour via smartphones we
need to limit the impact of these a-priori expectations and other biases on the
process of deriving digital phenotypes. Recently, we made the first step towards
deriving data-driven and unbiased phenotypes from multimodal smartphone
data. We showed by using a traditional measure of social functioning that this
digital phenotype captures several aspects of daily social functioning (chapter 4).
This phenotype was derived by using an unsupervised deep learning approach
(variational autoencoder\textsuperscript{26}) which translated the high dimensional and multimodal
smartphone data into a low dimensional latent space. The derived phenotype
of daily social functioning was found in this low dimensional latent space. The
premise of this approach is that the derived phenotypes are unbiased and not
driven by any domain knowledge and a-priori expectations about behaviour
in specific populations. However, it is important to emphasize the limitations
of these data-driven approaches. First of all, these data-driven, or so-called
unsupervised algorithms are able to detect nuanced biases\textsuperscript{27} in the data which
might translate back into the derived phenotypes. For example, without providing any demographical factors to the model, the derived phenotypes of daily social functioning were strongly associated with age and employment status. Given the relevance of these demographical factors to daily functioning this was not a major concern for our model. However, these results demonstrate the sensitivity of these data-driven models to detect the underlying biases that are present in the raw data collected by smartphones and caution is needed during the behavioural interpretation of these models. A second limitation of this approach is reproducibility of the phenotypes. In order to derive these phenotypes from a model a set of weights are needed that are optimised by training a model. Researchers should share these weights and model definitions in order for researcher to derive the same behavioural phenotype.

Figure 1 | Overview of the pre-processing procedure for smartphone-based location data. (A) Raw location data obtained by sampling the GPS sensor of a smartphone. (B) Context-enriched location data obtained by using a pre-processing procedure based on validated methods in the context of location data. In this example three recurrent stationary states were identified by using this method.
MONITOR DISEASE SEVERITY TO PREDICT ONSET AND RELAPSE OF DISEASE

The majority of medical disciplines utilize a measurement-based approach for the prevention and early detection of major medical events by monitoring measures that indicate the onset of such an event. For example, patients with hypertension and/or diabetes are regularly monitored to prevent further progression of the disease and major medical events that might lead to further complications and prolonged disabilities. These measurement-based approaches are proven to be effective in terms of disease progression and health economics\(^{28,29}\). However, in psychiatry treatment is often preceded by a major psychiatric crisis or even hospitalization and is followed by a long period of disability and slow recovery. The focus on acute care in psychiatry is due to a lack of valid temporal and objective measures of behaviour that relate to overall mental wellbeing. Several Behavioural changes in day-to-day functioning are identified as prodromal symptoms or early indicators of relapse for several neuropsychiatric disorders\(^{30}\). For instance, prior to the onset of psychosis in schizophrenia, patients tend to withdraw from social relations and show subtle deficits in cognition and bodily functioning\(^{31}\). Similar indicators for onset and relapse are identified in the majority of psychiatric disorders such as depression, bipolar and anxiety disorders\(^{30}\). A measurement-based approach is needed in psychiatry to continuously monitor these indicators in real-life and to detect the onset and/or relapse of disorders and thereby potentially improve patient outcomes\(^{32}\).

Smartphone-based monitoring has the potential to be the first step towards a more measurement-based approach in psychiatry by continuously monitoring several aspects of behaviour that are indicative for the onset or relapse of a psychiatric disorder. This approach could act as an early warning sign system for healthcare professionals to initiate interventions focused on impeding the progression or the onset of a disorder. With this, healthcare providers could monitor those with a predisposition for developing a psychiatric disorder or patients that are in remission so that they can act upon warning signs. For instance, it has been demonstrated that by monitoring communication and mobility patterns as registered by a smartphone, healthcare providers can monitor daily social functioning\(^{33}\). Deviations in this daily social functioning could subsequently be utilised to capture signs related to the onset or relapse of schizophrenia of which are both known to be associated with changes in social functioning\(^{31}\). A recent study demonstrated the feasibility of such a warning system in schizophrenia. This study showed anomalies in mobility and social behaviour as measured through smartphones two weeks prior to a relapse\(^{18}\). Similar results have been found in the context of bipolar disorder. By using a variety of different smartphone modali-
ties such as mobility patterns\textsuperscript{8–11}, smartphone usage\textsuperscript{12,13}, smartphone-based social behaviour\textsuperscript{10,12,14} and voice recordings\textsuperscript{12,34} researchers were able to predict manic and depressive states in bipolar disorder. These results together suggest that for schizophrenia and bipolar disorder a smartphone-based warning system has the potential to detect crucial changes in behaviour that might indicate the relapse or onset of a disorder. The prevention of relapse across the wide spectrum of psychiatric disorders is known to improve patient outcomes\textsuperscript{35}. In addition to the improvement patient outcomes, a measurement-based approach in psychiatry also has the potential to reduce the enormous economic burden of psychiatric disorders on society\textsuperscript{36}. This is especially the case if such a measurement-based approach is able to prevent the progression of psychiatric disorders. However, important to emphasize is that further research is needed to determine the efficacy of a smartphone-based approach in preventing the progression of these disorders.

In order to assess the efficacy of a smartphone-based monitoring approach in psychiatry, researchers need to (1) reproduce earlier findings and identify relevant behavioural indicators, (2) study the feasibility of this approach in other disorders such as, but limited to, depression, anxiety and Alzheimer’s disease, and (3) conduct randomized trials to assess the efficacy of such an approach in terms patients outcomes. In particular the latter is crucial for translating this now mainly academic tool for quantifying behaviour to clinical practice. To achieve this, researchers need to demonstrate through randomized trials that this approach is effective in reducing relapse rates and that it prevents the onset of disorders. Currently, the first randomized trails are being conducted in bipolar disorder\textsuperscript{37} and depression\textsuperscript{38} and aim to assess the efficacy of a smartphone-based approach in the prevention of disorders. For these trials to succeed it will be crucial that the involved researchers apply proper statistical approaches to analyse the longitudinal data.

**RECRUITMENT BIASES IN SMARTPHONE-BASED MONITORING**

The quantification of behaviour through smartphones addresses some of the major limitations that are inherent to traditional measures of behaviour (i.e. questionnaires, interviews and in-person assessments). However, it is also crucial to identify and discuss the limitations of this approach and to what extent these limitations are impeding the validity of the obtained results. The most notable limitation of this approach is related to multiple aspects of the recruitment of participants that are willing to collect data via their smartphone. The main aim of a good recruitment phase is to (1) find an adequate set of participants that repre-
sent the target population and (2) to obtain a reasonable sample size to perform groupwise comparisons. Systematic biases in the recruitment phase impedes the generalisability of the obtained results and could lead to misleading conclusions regarding the target population.\(^{39-42}\)

A recent clinical review that discussed the engagement of participants with mental health-related applications revealed an overall poor adherence\(^{43}\) to studies that utilize a smartphone-based approach\(^{44-46}\). These low-engagement numbers can be partly explained by the decreased motivation which is an often seen symptoms in many psychiatric disorders and contributes to poor treatment adherence.\(^{47}\) For example, a study that aimed to evaluate the engagement of schizophrenia patients with a smartphone application revealed that only 49% (\(n = 622\)) agreed to passively monitor their symptoms. They also showed that this percentage of patients that agreed to monitor their symptoms rapidly declined over time and that only a limited number of patients collected data over a longer period.\(^{48}\) In order to use this longitudinal data for inference regarding this populations, we need to identify the characteristics of those patients that adhere to the longitudinal monitoring of their symptoms and assess if non-adherence is related to a specific sub-population. To our knowledge, studies to date that utilize a smartphone-based approach to monitor behaviour often do not report the demographical or symptomatic characteristics of those participants that refuse to participate.

We used the data of the Psychiatric Ratings using Intermediate Stratified Markers (PRISM) project\(^{49}\) to assess if refusal of participation is associated with the severity of symptoms in schizophrenia (SZ) or Alzheimer’s disease (AD). The PRISM project utilized a smartphone application to passively monitor daily social functioning in SZ (\(n = 57\)), AD (\(n = 52\)) and age- and sex-matched controls (\(n = 59\)) and aimed to get a better understanding of the underlying biological mechanisms associated with the variations in social functioning. Here we show that the refusal of participation in SZ is associated with significantly higher positive symptoms (as measured by the positive and negative syndrome scale (PANSS)) in those patients (\(F(1,25) = 5.161, p = .032\), Figure 2) that refused to participate due to privacy reasons (\(n = 12\) vs \(n = 15\)). Given that delusions are a core feature of the positive symptoms scale, participation in passive behavioural monitoring might intensify these delusions and increases stress among these patients. The average positive symptoms score in SZ is 18.20\(^{51}\) and substantially higher than the average of 10.20 for those patients that agreed to collect data through their smartphones in PRISM project. These results suggest a recruitment bias related to the severity of SZ and limits the ability to generalize the results of this study to individuals diagnosed with SZ. If we take this bias into account, the generalization of the results from
this study should focus on those patients with a relative mild degree of positive symptoms.

This example of a recruitment bias related to the symptomatic characteristics of patients emphasizes the importance of the study design while using this novel and digital approach. In the PRISM project, the collection of behavioural data through smartphones was not mandatory for the participants. It is likely that this contributed to the bias in those SZ patients that agreed to participate. As a consequence of this the generalization of the obtained results is limited. If participation is not mandatory for these methods, researcher should monitor the recruitment of participants and make sure that particular demographical or symptomatic characteristics are included in the study. In chapters 2 to 5, smartphone data from SZ patients that were recruited via PRISM projects was analysed and the results generalised to the complete SZ population. With the aforementioned recruitment bias, these results are perhaps only valid for the SZ population with mild and below average positive symptoms. Additional studies with valid and generalizable samples of SZ patients are need to reproduce these findings to confirm the generalizability of our results.

Figure 2 | Comparison of the severity of symptoms between those who participated and those who did not participated in SZ patients. Post-hoc comparison using the Tukey HSD test indicated that the mean positive symptoms score for those who participated (M = 10.2, SD = 0.92) is significantly lower than those SZ patients that did not participate due to privacy reasons (M = 13.3, SD = 1.03).
LOCATION DATA

The use of smartphone-based location data is rapidly gaining traction in research as a tool to quantify behaviour. The rational of using smartphone-based location data in research is its ability to capture social and mobility related indicators of behaviour in a longitudinal manner\(^\text{20}\). For instance, this source of data can be used to derive indicators of distance travelled, time spent at home and the number of places visited. These indicators have been used to study community participation and withdrawal from social relations in several health related research topics\(^\text{52,53}\). For example, a study with 22 bipolar patients and 14 healthy controls showed that the variation in mobility is indicative for a depressive state in bipolar disorder\(^\text{8}\). In another study, researchers collected smartphone sensor data of 17 schizophrenic patients over a period of three months. With this data they showed that two weeks prior to a relapse the number of social-behavioural anomalies, as partly measured by location data, increased significantly. Collectively, these studies suggest that smartphone-based location data can be used to obtain exclusive insights into several different aspects of behaviour that are clinically relevant\(^\text{25}\).

To better understand smartphone-based location data it is important to emphasize the difference between traditional GPS data and smartphone-based location data. The latter is obtained by combining the GPS sensor in smartphones with nearby wireless network access point data and the triangulation between cellular phone towers\(^\text{54,55}\). Relative to traditional GPS data, the advantage of this approach is that the collection of data is less dependent on the signal strength of the GPS sensor. This signal strength is often influenced by external factors, this includes the weather and other factors related to the environment (i.e. large buildings and trees) that can obstruct the signal. This approach of combining different modalities to monitor location data is also needed to account for the relatively weaker GPS sensors that are embedded in contemporary smartphones. A by-product of this combination is that smartphone-based location data is relatively accurate in areas with many nearby wireless network access points and cellular phone towers (i.e. urban areas). Inherent to this is that smartphone-based location data is likely to be less accurate relative to traditional GPS in rural areas with limited to no availability of these two supporting modalities.

Despite that smartphone-based location data offers a unique opportunity to monitor mobility related behaviour in a longitudinal and passive manner, it is crucial to identify other factors that affect the quality of location data collected through smartphones. Factors that likely influence the quality of the data and often over seen in earlier research relate to demographical and socioeconomical characteristics of participants. For example, a recent study showed differences in data quality between devices and that caution is needed during the analysis of
Methodological and statistical perspectives on smartphone-based digital phenotyping

A potential explanation for this difference in quality are device characteristics that relate to the sensitivity and precision of the sensors embedded in smartphones. More sensitivity and precision are likely to be associated with increased consumer prices of smartphones and suggest a socioeconomical effect on the quality of the data collected through smartphones. This argument is partly supported by several studies which showed that the usage and adaptation of smartphones in daily life is associated with several demographical, cultural and socioeconomical factors. A potential side-effect of this low adaptation smartphones in daily life is that smartphones are for instance left at home while location data is passively monitored on the background. This particular example leads to the overestimation of the time spent at home while the individual is actually traveling and/or is visiting other locations. Research showed that the adaption and usage of smartphones in daily life is significantly associated with age, education, income and relationship status. Younger individuals with a high education and income tend to use their smartphone more frequently on a daily basis relative to their counterparts. These data also suggest that individuals with specific demographical characteristics tend to leave their smartphone at home more often and thereby affecting the validity of the location data collected. In chapters 3, 4 and 5 we used smartphone-based location data to assess social functioning in AD and SZ patients. Especially for AD and their age- and sex-matched controls, smartphone adaptation in day-to-day life is relatively low compared to the SZ group due to age. Due to this potential low adaption, estimates of for example the time spent at home might be overestimated in this group and potentially represent somewhat skewed estimates of reality. To account for this effect, we need to identify the relevant demographical characteristics and adjust for these effects in the analysis of the data. An additional solution for this limitation is to combine accelerometer and usage log data to identify longer periods were the smartphone is not being used which might be indicative that the smartphone is left at home for example. With this, researchers can exclude specific periods of the data collection for any further analysis and with that increase the validity of the derived behavioural phenotypes.

CONCLUSION

Smartphone-based phenotyping is an important and promising step towards a quantitative and more objective assessments of behaviour in humans. The development of novel behavioural features based on digital data has the potential to enhance our understanding of the variations in behaviour and their underlying biological mechanisms. For neuropsychiatric disorders, this is the first step towards
a scalable and more objective and quantitative measure of behaviour, which will be a critical step forward to improve our understanding of mental disorders in particular and human behaviour in general. By employing this novel digital approach in the context of mental disorders we can more effectively study disease progression, treatment efficacy, and identify early indicators of disease that allow for the prediction of disease onset, relapse and remission of these disorders. In addition, researcher argue that the quantitative and objective characteristics of this novel digital approach might contribute to the identification of the biological pathways involved in mental disorders\textsuperscript{49}. The identification of the biological pathways involved in these disorders could enhance the development of specific treatments that target these pathways and hopefully improve patient outcomes. In chapters 4 and 5 we showed, to our knowledge for the first time, that social behaviour phenotypes derived by using this digital approach are successfully associated with several biological pathways known to be involved in social functioning. The strategies outlined in these chapters provide efficient tools to further study the characteristics of daily social functioning and other phenotypes across neuropsychiatric disorders and to further study their neurobiological underpinnings.

One of the most often used arguments in favour of smartphone-based phenotyping is its ability to quantify behaviour in an objective manner. However, this argument trivializes the most important limitation of this novel and digital approach. Namely, the biases in the observed data and derived phenotypes that relate to demographical and clinical characteristics of the sample population, which is similar to the biases in traditional assessments of behaviour. These biases might impact the validity of the derived phenotypes and the conclusions regarding the behaviour in the target population. A more nuanced and alternative argument would be to state that smartphone-based phenotyping is more objective method to quantify behaviour relative to its traditional counterparts (i.e. questionnaires and self-reports). To utilize the full potential of this smartphone-based approach we need to identify those demographical and clinical characteristics that are relevant and adjust for them accordingly in the analysis of the data. Thus, digital phenotyping provides a novel and promising avenue to expand our understanding of human behaviour in health and disease. Future studies should focus on the validation of digital behavioural endpoints, study their value as a clinical biomarker in, for example, treatment efficacy studies, as well as an early indicator in the development of neuropsychiatric disorders, and/or as early warning signal for relapse, among others. In light of these promises, future actions should also include the early engagement with regulators, improving our understanding on the general uptake of these digital methodologies, and addressing some of the technological challenges that come with the general application of smartphone-based data sampling.
REFERENCES


35. Ho, F. Y. Y., Yeung, W. F., Ng, T. H. Y. & Chan, C. S. The Efficacy and Cost-Effectiveness of Stepped Care Prevention and Treatment for Depressive and/or Anxiety Disorders:


