



Published in final edited form as:

*J Psychiatr Brain Sci.* 2020 ; 5: . doi:10.20900/jpbs.20200010.

## Grant Report on SCH: Personalized Depression Treatment Supported by Mobile Sensor Analytics

Jayesh Kamath<sup>1,\*</sup>, Jinbo Bi<sup>2</sup>, Alexander Russell<sup>2</sup>, Bing Wang<sup>2,\*</sup>

<sup>1</sup>Psychiatry Department, University of Connecticut Health Center, Farmington, CT 06030, USA

<sup>2</sup>Computer Science & Engineering Department, University of Connecticut, Storrs, CT 06269, USA

### Abstract

We report on the newly started project “SCH: Personalized Depression Treatment Supported by Mobile Sensor Analytics”. The current best practice guidelines for treating depression call for close monitoring of patients, and periodically adjusting treatment as needed. This project will advance personalized depression treatment by developing a system, DepWatch, that leverages mobile health technologies and machine learning tools. The objective of DepWatch is to assist clinicians with their decision making process in the management of depression. The project comprises two studies. Phase I collects sensory data and other data, e.g., clinical data, ecological momentary assessments (EMA), tolerability and safety data from 250 adult participants with unstable depression symptomatology initiating depression treatment. The data thus collected will be used to develop and validate assessment and prediction models, which will be incorporated into DepWatch system. In Phase II, three clinicians will use DepWatch to support their clinical decision making process. A total of 128 participants under treatment by the three participating clinicians will be recruited for the study. A number of new machine learning techniques will be developed.

### Keywords

depression; personalized depression treatment; mobile sensing; data analytics; machine learning

### INTRODUCTION

Depression is a complex, heterogeneous, severely debilitating and chronic illness. It affects more than 264 million people worldwide, contributing to significant number of deaths by suicide every year [1]. The prevalence rate of depression varies across the world from 3% in Japan to 16.9% in the United States, with most countries reporting between 8% and 12% [2]. Due to its high lifetime prevalence and its effects on daily function and its mortality [3], depression is expected to become the world’s largest medical burden of disease by 2020 [4].

Licensee Hapres, London, United Kingdom. This is an open access article distributed under the terms and conditions of [Creative Commons Attribution 4.0 International License](#).

\*Correspondence: Jayesh Kamath, jkamath@uchc.edu; Tel.: +1-860-679-6727; Bing Wang, bing@uconn.edu; Tel.: +1-860-486-0582.

#### CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

The goal in treating depression is achieving symptom remission and full functional recovery [5]. Similar to other fields of medicine, there has been a strong impetus in the field of psychiatry to personalize depression treatment, i.e., quickly identify the best treatment for a depressed individual while minimizing side effects in the clinical setting. However, despite decades of research, finding the perfect treatment for a patient has been elusive—very few clinical characteristics, biomarkers, or genetic variations have been identified that can reliably predict differential effectiveness or adverse effects of specific depression treatments [6–8]. Clinical guidelines such as the Canadian Network for Mood and Anxiety Treatments (CANMAT) Clinical Guidelines provide recommendations for treatments for major depression [9]. However, the guidelines acknowledge significant limitations of evidence-base, lack of comparative trials, heterogeneity of patient populations, and several other factors such as drug-drug interactions, pharmacodynamics and pharmacokinetic factors [9]. In the context of this limited evidence-base, guidelines recommend tailoring treatments to the patient, i.e., accounting for diverse factors in clinical setting and adjusting treatments as needed [10,11]. This “trial and error” treatment strategy can have serious detrimental consequences in clinical setting, e.g., negative impact on patient-provider therapeutic alliance, patients dropping out of treatment [12]. It has been reported that only about 35% of patients will remit upon initial treatment in a given episode [13]. On the positive side, for the patients who did not remit after the first treatment, up to 40% experience significant improvement after switching to an alternative treatment, adding a second medication, or adding psychotherapy provided they pursue these options [13]. The reality, however, is that patients typically drop out of treatment if initial treatments fail or they experience side effects: After starting antidepressant treatment, nearly half make no follow-up visits and only one quarter return often enough to pursue additional treatment options [12].

To optimize patient care outcome, it is highly desirable to predict whether a patient will eventually respond to treatment early in the course of the treatment. Recently, a particularly exciting discovery is that early improvement or lack thereof in the first two weeks after initiation of an antidepressant treatment is a good predictor of later full response to treatment [14–21]. This finding highlights the value of early assessment and close follow-up with patients, particularly important for those who have shown no improvement. In existing studies, close monitoring has been achieved using physician-administered follow-ups or patient self-administered questionnaires. While feasible in well-executed and well-funded studies, this is difficult to carry out in clinical settings for a number of reasons: Firstly, frequent follow-ups by clinicians is difficult due to the significant lack of trained professionals—in the United States, the ratio is 14.5 psychiatrists per 100,000; in developing countries, the ratio is more than ten times lower [22]. Secondly, patient self-administered questionnaires are burdensome, and responses to these questionnaires are often subjective and limited to recall bias. Commonly used depression questionnaires in the clinical setting such as the Patient Health Questionnaire (PHQ-9) ask patients to report their emotional status and other symptoms over preceding two week period [23]. Patient’s responses to these questions are frequently colored by their emotional status at the moment when they complete these questionnaires. Thus, responses to these questionnaires fail to capture objective data including day-to-day variations in patient’s depression status, its behavioral manifestations, and its impact on patient functioning [24]. Impairment in functioning is a

critical depression criterion according to the Diagnostic and Statistical Manual of Mental Disorders (DSM V) and an important determinant of depression severity [25]. Depression questionnaires also fail to reliably and objectively capture change in depression severity, especially changes that occur in response to treatment initiation [24].

Smartphones, with their multiple sensors and increasingly advanced computing capabilities, have the potential to serve as “human sensors” by monitoring users in real time for changes in their behavioral patterns. Several reports provide evidence of feasibility and potential efficacy of using smartphone based data for clinical inferences in the management of affective disorders, primarily depression and bipolar disorder (by our team and other research groups, see reviews in [26–32]). Specifically, in the LifeRhythm Project, a 4-year project funded by the National Science Foundation, our group conducted a two-phase study in college age participants with depression (in comparison with a control group). Our results demonstrated that sensory data collected from mobile and wearable devices—without any user interaction—can provide critical information that correlates with depression symptoms, and can be used to automatically detect depression [33–38]. Specifically, in Phase I of the project, we developed a smartphone application, called LifeRhythm app, to passively collect sensory data (location, activity, social interaction) for both Android and iPhones, the two predominant smartphone platforms. We further developed feature extraction techniques to extract behavioral features from the sensory data as correlates of depression symptomatology, and machine-learning models to predict self-report questionnaire scores and depression status (i.e., whether one is depressed or not). These techniques and prediction models were then validated and refined in Phase II of the study. In addition, in Phase II, we further explored using wristbands (Fitbit devices), in addition to smartphones, for characterizing behavioral features that are correlated with depression.

The current project builds on the insights and experiences we gained from the LifeRhythm project. The goal is to develop a system, DepWatch, that leverages mobile health technologies as well as machine learning tools to closely monitor patients’ depression symptoms and assist clinicians’ decision making. Specifically, the system aims to provide timely assessment of depression symptoms through mobile data analytics. Such timely assessment of depression can help clinicians personalize treatment by (i) identifying patients who are failing treatments early, and (ii) assisting them to take necessary actions before patients drop out of treatment. The project comprises two studies. The focus of Phase I study is to extend the LifeRhythm app to include self-report questionnaires on mood, anxiety, medication adherence, and medication tolerability and safety. The self-reports and the passively collected sensory data will then be used to develop, evaluate and cross validate machine learning models for assessing depressive symptoms and predicting patients’ response to treatment in the future e.g., in the two, four or six weeks after treatment initiation. At the end of Phase I, we will develop a web portal that will assist clinicians’ decision making (e.g., to continue the current treatment or to change the treatment) by incorporating the machine learning models. The web portal will be developed in collaboration with clinicians, by closely seeking their input and feedback through focus groups. In Phase II study, the web portal will be used by clinicians to evaluate its usability and efficacy. We hypothesize that DepWatch system will be able to predict response/non-response to depression treatments by capturing change in behavioral patterns as it relates to

changes in patient's depression severity, which will be useful in clinicians' decision making process.

This is a four-year project, currently in its first year. It is a joint interdisciplinary project between the University of Connecticut (UConn) and University of Connecticut Health Center (UCHC). It is led by Dr. Bing Wang, Dr. Jinbo Bi and Dr. Alexander Russell from UConn and Dr. Jayesh Kamath from UCHC. The project is broadly related to the studies that use smartphones and wearable devices to monitor, manage and assist the treatment of affective disorders (see reviews in [26–32]). As an example, MONARCA I and II trials feature patient self-monitoring using both objective sensory data and subjective self-assessment on smartphones; in addition, the data can be visualized on a web portal that can be accessed by both patients and clinicians [39,40]. The studies in [41,42] explore context-sensitive intervention delivery via smartphones to people with depressive symptoms to provide them in-situ support. Our study focuses on providing clinicians a decision support system that helps them to evaluate and adjust depression treatment, leveraging mobile sensor analytics.

### Aims of the Grant

**Aim 1**—Develop DepWatch system. DepWatch collects sensory data passively from smartphones and wristbands, without any user interaction, and uses simple user-friendly interfaces to collect ecological momentary assessments, medication adherence and safety related data from patients. The collected data will be fed into machine learning models to be developed in the project to provide weekly assessment of patient symptom levels and predict the the likelihood of whether a patient will respond in the next several weeks. The assessment and prediction results are presented using a graphic interface to clinicians to help clinicians make treatment decisions.

**Aim 2**—Extract higher-level structures from sensing data. While low-level features directly extracted from sensing data (e.g., entropy of locations, the amount of time spent at home) are correlated with depression, the correlation tends to be low. This project will identify features that provide semantic information or higher-level structures on user behaviors, which are potentially more correlated with depression symptoms. Specifically, the features to be explored include environment context (e.g., restaurant, movie theater, work place, gym) and routines that a user follows regularly that can describe the structures of the user's life.

**Aim 3**—Develop new machine learning algorithms. The focus will be on (i) longitudinal prediction, which is necessary since response to a medication may only be assessable after a period of time and the depression symptom change of a patient at the current time can depend on the features in the past, and (ii) multi-task feature learning for the challenging setting with large-scale heterogeneous data.

## BACKGROUND AND HIGH-LEVEL APPROACH

### Background

The current practice for treating depression is a “trial and error” approach that commences a course of medication and requests that patients to return to evaluate the effects later on. In the absence of emergency situations i.e., a patient calls due to severe side effects or suicidal ideation, the interval between two visits is typically 4 weeks. While the guidelines may recommend shorter intervals, more frequent follow-ups are difficult in clinical setting due to shortage of resources, particularly limited availability of psychiatrists. The above practice has several drawbacks. First, the clinician has no way of monitoring symptom evolution of the patient until the next follow-up visit. Even at that time, the symptom reports are based on questionnaires, which are subjective and known to suffer from recall bias [24]. The delayed awareness of a patient’s status may lead to delayed change in treatment strategies, potentially leaving a patient to more impairments associated with depression, higher risk of suicide, and higher likelihood of discontinuing treatment prematurely. In addition, this practice does not leverage the evidence that the presence or absence of early improvement (within two weeks after starting a new treatment strategy) can predict the treatment outcome (full response or not) in the future.

There is an urgent need of a system that (i) provides objective measurements and assessments to clinicians frequently e.g., on a weekly basis, with minimal burden to patients; and (ii) predicts treatment outcome (i.e., whether continuing the current treatment is likely to lead to remission later on) based directly on brief subjective report and objective behavior data so as to help clinicians best make treatment decisions.

### High-Level Approach

We will develop DepWatch, an automatic data collection and analytics system, to support and inform clinicians’ decision making for depression treatment. Development of machine learning models for the DepWatch system uses both QIDS-SR (Quick Inventory of Depressive Symptomatology-self report) scores [43] and clinician assessment as ground truth. However, a higher emphasis is placed on the clinician assessment which includes review of weekly QIDS scores and participant interview on a monthly basis or twice per month if clinically indicated. The QIDS has been validated in clinical settings. It provides more details on depression symptoms than another widely used self-report questionnaire, PHQ-9 [23]. Specifically, QIDS separates certain symptoms into individual items, e.g., insomnia into initial, middle, and terminal insomnia; appetite into low vs high appetite; weight into low vs high weight; and psychomotor symptoms into agitation vs retardation. PHQ-9 combines each of these sets of symptoms into one item. For our study purpose, i.e., to develop prediction models on the various depressive symptoms, it is critical to separate these symptoms. At the same time, QIDS is still not too burdensome and reflects the DSM V diagnostic criteria [43].

Figure 1 illustrates how DepWatch works: It collects patient data, encrypts the data, and sends it to a secure server. At the server side, the data will be preprocessed and analyzed with machine learning algorithms (that have been previously trained) to directly assess the

patient's current status. The results can be visualized by the clinicians using a graphic interface. More importantly, the results are primarily based on objective data gathered by DepWatch, capturing changes in behavioral patterns as it relates to changes in patient's depression status. A clinician then logs into DepWatch regularly to review the status of his/her patients e.g., on a weekly or biweekly basis, and can leverage these analytics to help decide whether the current treatment plan needs to be changed. In this case, the clinician will ask the patient to come back for a more detailed evaluation or connect with patients by phone to discuss ongoing treatments. In addition, the clinicians can receive automated alerts if their patients' exhibit behavior that suggests a significant negative change in symptoms or questionnaire responses that indicate suicidal ideation or cessation of antidepressant use, e.g. due to side effects or due to lack of efficacy.

Through DepWatch, a clinician can monitor a patient on a continuous basis, and change treatment plans in a timely manner to avoid adverse impacts on patients due to delayed adjustment of treatment plans. Since only the patients that need to be evaluated earlier, as identified by DepWatch, will be called back (the rest come back at their regularly scheduled visits), the clinic resources are used more efficiently. When a patient comes back for a follow-up visit, the clinician already has detailed information on the patient's mental health status to make better decisions. As a result, the treatment for patients is more personalized. From a patient's point of view, DepWatch allows a patient to be more engaged in the treatment process. As a result, we hope that patients will be more willing to continue their treatment and seek alternate medications if necessary.

## STUDY DESIGN AND DATA COLLECTION

The study is organized around two phases (see the timeline of the project in Table 1). The two primary goals of Phase I are (i) data collection and (ii) development, training, and cross-validation of the machine learning models, which will be incorporated in DepWatch system. In Phase II, we will arrange for clinicians to actively use the DepWatch system developed in Phase I to evaluate usability and efficacy. Subject recruitment of both phases will proceed through IRB-approved announcements and procedures. In this section, we briefly describe the study procedures; more details on the technical challenges are deferred to later sections.

Phase I study will recruit a total of 250 participants from several UConn Health outpatient clinics and from surrounding communities. Participants will meet the following inclusion/exclusion criteria: age 18 and above, have unstable depression of at least moderate severity as defined by a score of  $\geq 11$  on the QIDS questionnaire [24], initiating a pharmacological treatment (monotherapy or adjunctive treatment) for depression, no current or past diagnosis of bipolar disorder or primary psychotic disorders such as schizophrenia, and no clinically significant medical, psychiatric, or substance use comorbidities that may adversely affect participant's study participation and/or affect their adherence to study protocol e.g., significant cognitive deficits, clinically significant substance use disorder within one month of study enrollment. The participation of each participant will last up to 12 weeks. All subjects meeting the study eligibility criteria will complete the QIDS at baseline. At the baseline visit, i.e., at the time of enrollment, participants' demographic and clinical information will be collected e.g., medical and psychiatric comorbid conditions, past



treatment history/medication trials. Participants will be asked to download and use the LifeRhythm app on their smartphones. The LifeRhythm app was developed in a prior project to passively collect objective sensory data, and is extended in this project to collect brief questionnaires, including QIDS (weekly), mood and anxiety (daily), medication adherence, medication tolerability and safety (weekly). Participants will also be provided a Fitbit wristband (if they do not have one) to collect specific physiological information, e.g., sleep, heart rate, activity. Monetary incentives/compensation will be provided to the participants for their participation and adherence to study procedures. The study clinician will follow up with a patient once per month (or twice per month if clinically indicated) to correlate QIDS reports with patient's verbal reports through phone call or secure video call. The study clinician will closely monitor patients for worsening of depression symptoms, and will coordinate communication with the patients' regular clinicians if significant worsening of symptoms with emergence of active suicidal thoughts with intent and plan noted based on study-related assessments.

In the Phase II study, three clinicians will use DepWatch to support their clinical decision making process. A total of 128 participants (aged 18 and above, each treated by one of the three clinicians) will be recruited for this study phase. The inclusion/exclusion criteria will be the same as for Phase I. For half of the participants, their clinicians will use DepWatch system, i.e., they will be prompted to review patient data/assessments gathered by the study team, and use the prediction model to assist their clinical decision when appropriate. The other half of the participants will serve as the control group; their clinicians will not use DepWatch system during their treatment. Changes in treatments initiated by the clinicians in response to the reports provided to them and overall depression outcomes for both groups will be assessed.

## MACHINE LEARNING MODELS AND DATA ANALYTICS

In Phase I study, we will use the collected data to train a family of new machine learning models. These machine learning are in two broad categories: (i) time-series models for predicting QIDS scores and clinical depression, e.g., remission, mild depression, severe depression, and (ii) direct prediction of changing depression status, e.g., significant improvement, mild improvement, no significant change, etc. The first category of models are trained directly against collected QIDS scores and clinical diagnosis, and aim to predict these via a variety of collected data e.g., sensory data, demographic information, past treatment history, mood/anxiety data. The second category of models are geared towards directly detecting significant changes in depression during the few weeks after initiation of treatment, one of the principal goals of this project. While changes in depression can be detected based on the prediction of QIDS or clinical diagnosis i.e., through the first category of models described above, we anticipate that higher reliability can be achieved by directly training our models against the associated "categorical" clinical ground truth. Specifically, we will perform two types of predictions: (i) determining whether a patient's status is worsening, remains to be about the same, has improved, or has improved significantly, and similarly determining a patient's status in specific symptoms e.g., interests, sleep, psychomotor; in both cases we also wish to identify the specific features relevant to the

prediction, and (ii) predicting longer-term trajectory without changes in treatment, i.e., the likelihood of improvement in the next two to four weeks.

One significant challenge in developing the above machine learning models is that the data are complex and heterogeneous, involving diverse types of objective sensory data (e.g., location, activity, social interaction) collected on different platforms (e.g., Android, iPhones, Fitbit), subjective self-reports (e.g., mood, anxiety, medicine adherence), and clinical data (e.g., history of medication). In addition, the prediction targets are not directly compatible: QIDS score is a numerical outcome, while depression severity is a categorical outcome. We envision that multi-task feature learning (MTFL) methods are necessary and beneficial to our setting. Specifically, we can use MTFL methods to construct sparse mappings from the heterogeneous data to the multiple targets of predicting QIDS scores and depression severity categories. Our prior study [36] developed a MTFL method that jointly builds inference models for related tasks of different types, including classification and regression tasks, based on sensory data, which will be expanded to include even more heterogeneous data sources in this study.

Another significant challenge in our study is that the data is longitudinal, and the data records collected at different points in time are not independently and identically distributed. Specifically, because response to a medication may be visible or assessable only after a reasonable accumulation of time, the depression symptom change of a patient at the current time point could depend on not only the data records of the current time period e.g., the current week, but also on the data collected in the past e.g., the past week, two weeks, or even longer periods of time. Building predictive models using longitudinal records that are not independently and identically distributed is challenging. We plan to leverage statistical methods to correct for such samples. In an early study [44], we applied a variant of such methods to study the fMRI data from patients of Alzheimer's Disease using a series of historical fMRI images that were taken at 24, 18, 12, 6, and 3 months in the past. We envision such an approach will be promising for this study.

The accuracy of the above machine learning models depends heavily on (i) data quality and sample size, and (ii) feature extraction. In our prior LifeRhythm study, we have found that missing sensory data is a prevalent and severe problem [35]. The problem will be further amplified in the current project due to the more diverse data that will be collected. The missing data problem becomes more challenging to address if certain subjects miss an entire view of data (e.g., a self-report, an entire day or multiple days of location data) rather than sporadic absence of a few measurements. Classic multiple imputations or matrix completion methods are clearly ineffective here because no information in the specific data source can be used to impute data for such samples. The commonly-used strategy of simply removing samples with a missing view or missing variables can dramatically reduce sample size, thus diminishing the statistical power of subsequent analyses [35]. We plan to leverage approaches based on deep learning models for data imputation. Specifically, we envision a promising direction is generative adversarial network (GAN) [45], which learns to generate new data with the same statistics as the training set. In addition to data imputation, we will further develop robust feature extraction techniques that provide features to the machine learning models. In earlier work [34–38], we have explored using various features from



location and activity data e.g., entropy of locations, the amount of time spent at home, circadian movement, similarity in location across days. While these features are correlated with depression, the correlation tends to be relatively low. In this study, we will identify features that provide semantic information or higher-level structures, which may be more correlated with depression. One direction is to identify environment context, e.g., restaurant, movie theater, work place, shopping plaza, outdoors, gym, social places, which are important in correlating with depression symptoms. Another direction is identifying higher-level routines that a user follows regularly and can describe the structures of the user's life.

## SYSTEM DEVELOPMENT AND EVALUATION

The DepWatch system contains two main components: (i) a data collection system that collects both objective sensory data and subjective patient self-reports to a secure server, and (ii) a web portal that visualizes patient information to clinicians to help their decision making. In the data collection system, the LifeRhythm app will be deployed on participants' phones to passively collect diverse sensory data (location, activity, social interaction) and brief self-report questionnaires, such as the daily cognitive assessment and weekly assessment of medication adherence, tolerability and safety. The self-report questionnaires use easy-to-use graphic interface and participants are sent notification to fill in the questionnaires at the due date directly on his/her phone. The data collected by the app will be encrypted at the phone, and transmitted to the secure server when the phone is connected to a WiFi network. Each participant is assigned a random ID. The collected data is associated with the random ID, instead of the real identity of the user. The Fitbit data is collected (after user authentication during the informed consent) using the Fitbit Subscription API provided by Fitbit to the secure server.

The web portal will display three main categories of information. The first category is the information that has been gathered, including behavioral features (e.g., the amount of time spent at home each day, the amount of time in bed each day), mood (e.g., mood category each day), tolerability and safety in a specified time period (i.e., the past 2 weeks). The goal is to present a patient's information to the clinician in an easily accessible form. The second category presents evolution of depression symptom (e.g., improving or worsening) in the general status and specific symptoms (e.g., interests, sleep, psychomotor) for a specified time period (e.g., in the past four weeks, relative to the baseline, or relative to the previous week). The third category is the prediction results, i.e., the likelihood of whether a patient will respond in four weeks with current treatment. Note that the results for the latter two categories are obtained from our machine learning models; they do not use QIDS reports at all (indeed QIDS reports are only collected for training and cross-validation the machine learning models; patients do not need to fill in QIDS when DepWatch is used in practice). The web portal needs to be designed so that it can be easily incorporated into clinical practice and be useful to clinicians' decision making. We will design it in close collaboration with clinicians, seeking their input and feedback through focus groups, e.g., what are the most important information to display, how to display the information, etc. The design of the web portal will go through an iterative process, with multiple rounds of design and refinement. In Phase II study, three clinicians will use the web portal to view information about their patients who elect to participate in the study as well as the prediction results. We

will encourage them to send us comments while using the system, and will improve the system based on their comments. Especially, we will seek their feedback and thought process when their clinical judgement differs from the results presented by the system. In the middle and at the end of Phase II study, we will ask the clinicians to evaluate the usability, accuracy, and efficacy of the system, which will be used for improving the system in the future.

Note that DepWatch system serves primarily as a decision support system for clinicians. It does not serve as a tool for patients self-monitoring. As a result, the design in the LifeRhythm app mainly focuses on providing easy-to-use interfaces for patients to enter the questionnaire information at the due date. It does not provide interfaces for patients to keep track of their past input. The web portal will only be used by the clinicians. Extending the system to provide information to patients is left as future work.

## CURRENT STATUS

The project started in August 2019. We have completed developing the data collection system. Figure 2 shows the two daily questionnaires (on mood and anxiety) and three weekly questionnaires (on safety, tolerability, medication adherence) that have been developed in the LifeRhythm app. The app runs in the background, passively collecting sensory data. When it is the time to fill in the questionnaires, notifications will be sent to the users to remind them to fill them in. We stated recruiting participants for Phase I study in January 2020. As the data are being collected, we will analyze the data and develop, train and cross validate machine learning models to provide symptom assessment and predict the trajectory of response.

## CONCLUSIONS

Current treatment of depression lacks strong evidence-base, objective and timely assessments, and treatment biomarkers. DepWatch, a mHealth system developed by our team, captures behavioral patterns as a correlate of depression status. The ongoing project investigates its utility to predict response/non-response to depression treatment in real-time. DepWatch will provide clinicians truly objective data on their patients depression status. Furthermore, such data provided in real-time will serve as a behavioral biomarker helping clinicians make critical treatment decisions. Development of DepWatch is a vital step towards personalized and patient-centric depression care.

## Acknowledgments

### FUNDING

This research was funded by the National Institute on Mental Health: grant number R01MH119678-01.

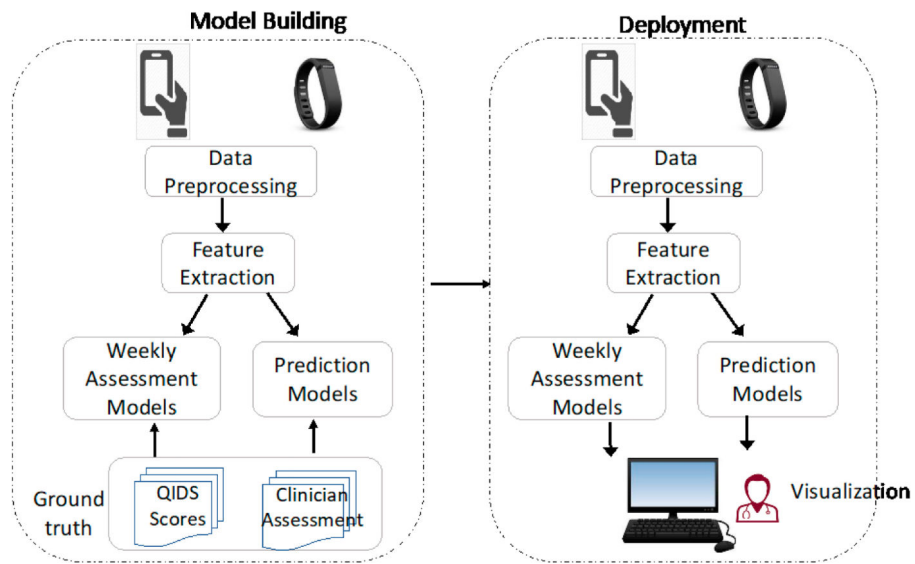
## REFERENCES

1. World Health Organization (WHO) 2020 Available from: <https://www.who.int/news-room/fact-sheets/detail/depression>. Accessed 2020 Apr 26.

2. Andrade L, Caraveo-Anduaga J, Berglund P, Bijl R, DeGraaf R, Keller M, et al. The epidemiology of major depressive episodes: results from the international consortium of psychiatric epidemiology (ICPE) surveys. *Int J Methods Psychiatr Res.* 2003;12(1):3–21. doi: 10.1002/mpr.138 [PubMed: 12830306]
3. Judd LL, Akiskal HS, Zeller PJ, Paulus M, Leon AC, Maser JD, et al. Psychosocial disability during the long-term course of unipolar major depressive disorder. *Arch Gen Psychiatry.* 2000;57:375–80. [PubMed: 10768699]
4. World Health Organization. 2013 Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013–2020. Available from: [http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236_eng.pdf). Accessed 2020 Apr 26.
5. Nutt D, Davidson J, Gelenberg A, Higuchi T, Kanba S, Karamustafalioglu O, et al. International consensus statement on major depressive disorder. *J Clin Psychiatry.* 2010;71(suppl E1):e08.
6. Kemp A, Gordon E, Rush A, Williams L. Improving the prediction of treatment response in depression: Integration of clinical, cognitive, psychophysiological, neuroimaging, and genetic measures. *CNS Spectr* 2008;13(12):1066–86. [PubMed: 19179943]
7. Simon GE, Perlis RH. Personalized medicine for depression: Can we match patients with treatments? *Am J Psychiatry.* 2010;167(12):1445–55. [PubMed: 20843873]
8. Cohen ZD, DeRubeis RJ. Treatment selection in depression. *Annu Rev Clin Psychol.* 2018;14(15):209–36. [PubMed: 29494258]
9. Kennedy SH, Lam RW, McIntyre RS, Tourjman SV, Bhat V, Blier P, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder: Section 3. Pharmacological Treatments. *Can J Psychiatry.* 2016;61(9):540–60. [PubMed: 27486148]
10. Morris DW, Touns M, Trivedi MH. Measurement-based care in the treatment of clinical depression. *Am J Psychiatry* 2015;172(10):1004–13. [PubMed: 26315978]
11. Fortney JC, Unutzer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement-based care. *Psychiatr Serv* 2017;68(2):179–88. [PubMed: 27582237]
12. Simon GE, Korff MV, Rutter CM, Peterson DA. Treatment process and outcomes for managed care patients receiving new antidepressant prescriptions from psychiatrists and primary care physicians. *Arch Gen Psychiatry.* 2001;58(4):395–401. [PubMed: 11296101]
13. Rush A, Trivedi M, Wisniewski S, Nierenberg A, Stewart J, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR\*D report. *Am J Psychiatry.* 2006;163(11):1905–17. [PubMed: 17074942]
14. Nierenberg A, McLean N, Alpert J, Worthington J, Rosenbaum J, Fava M. Early nonresponse to fluoxetine as a predictor of poor 8-week outcome. *Am J Psychiatry.* 1995;152(10):1500–3. [PubMed: 7573590]
15. Szegedi A, Müller M, Angheliescu I, Klawe C, Kohnen R, Benkert O. Early improvement under mirtazapine and paroxetine predicts later stable response and remission with high sensitivity in patients with major depression. *J Clin Psychiatry.* 2003;64(4):413–20. [PubMed: 12716243]
16. Szegedi A, Jansen W, van Willigenburg A, van der Meulen E, Stassen H, Thase M. Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: a meta-analysis including 6562 patients. *J Clin Psychiatry.* 2009;70(3):344–53. [PubMed: 19254516]
17. Henkel V, Seumüller F, Obermeier M, Adli M, Bauer M, Mundt C, et al. Does early improvement triggered by antidepressants predict response or remission? analysis of data from a naturalistic study on a large sample of inpatients with major depression. *J Affect Disord.* 2009;115(3):439–49. [PubMed: 19027961]
18. Kemp D, Ganocy S, Brecher M, Carlson BX, Edwards S, Eudicone JM, et al. Clinical value of early partial symptomatic improvement in the prediction of response and remission during short-term treatment trials in 3369 subjects with bipolar i or ii depression. *J Affect Disord.* 2011;130(1–2):171–9. [PubMed: 21071096]
19. Muzina DJ, Chambers JS, Camacho TA, Eudicone JM, Forbes RA, Berman RM, et al. Adjunctive aripiprazole for depression: Predictive value of early assessment. *Am J Manag Care.* 2011;17(12):793–801. [PubMed: 22216750]

20. Joel I, Begley A, Mulsant B, Lenze E, Mazumdar S, Dew M, et al. Dynamic prediction of treatment response in late-life depression. *Am J Geriatr Psychiatry*. 2014;22(2):167–76. [PubMed: 23567441]
21. Katzman M, Nierenberg A, Wajsbrot D, Meier E, Prieto R, Pappadopulos E, et al. Speed of improvement in symptoms of depression with desvenlafaxine 50 mg and 100 mg compared with placebo in patients with major depressive disorder. *J Clin Psychopharmacol*. 2017;37(5):555–61. [PubMed: 28817491]
22. Health at a Glance 2011: OECD Indicators Available from: <https://www.oecd.org/els/health-systems/49105858.pdf>. Accessed 2020 Apr 26.
23. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Int Med*. 2001;16(9):606–13.
24. Robinson J, Khan N, Fusco L, Malpass A, Lewis G, Dowrick C. Why are there discrepancies between depressed patients' Global Rating of Change and scores on the Patient Health Questionnaire depression module? A qualitative study of primary care in England. *BMJ Open*. 2017;7:e014519. doi: 10.1136/bmjopen-2016-014519
25. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington (VA, USA): American Psychiatric Association; 2013.
26. Gravenhorst F, Muaremi A, Bardram J, Grünerbl A, Mayora O, Wurzer G, et al. Mobile phones as medical devices in mental disorder treatment: an overview. *Pers Ubiquit Comput*. 2015;19(2):335–53.
27. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-Based Monitoring of Objective and Subjective Data in Affective Disorders: Where Are We and Where Are We Going? *Systematic Review*. *J Med Internet Res*. 2017;19(7):e262 [PubMed: 28739561]
28. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. *Int J Bipolar Disord*. 2018;6(1):6. [PubMed: 29359252]
29. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations Between Objective Behavioral Features Collected From Mobile and Wearable Devices and Depressive Mood Symptoms in Patients With Affective Disorders: Systematic Review. *JMIR Mhealth Uhealth*. 2018;6(8):e165. [PubMed: 30104184]
30. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med*. 2019;49:1426–48. [PubMed: 30744717]
31. Faurholt-Jepsen M, Geddes JR, Goodwin GM, Bauer M, Duffy A, Kessing LV, et al. Reporting guidelines on remotely collected electronic mood data in mood disorder (eMOOD)—recommendations. *Transl Psychiatry*. 2019;9:162. [PubMed: 31175283]
32. Razavi R, Gharipour A, Gharipour M. Depression screening using mobile phone usage metadata: a machine learning approach. *J Am Med Inform Assoc*. 2020;27(4):522–30. [PubMed: 31977041]
33. Farhan AA, Lu J, Bi J, Russell A, Wang B, Bamis A. Multi-view Bi-Clustering to Identify Smartphone Sensing Features Indicative of Depression. In: *Proceedings of IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*; 2016 Jun 27–29; Washington, DC, USA.
34. Farhan AA, Yue C, Morillo R, Ware S, Lu J, Bi J, et al. Behavior vs. Introspection: Refining prediction of clinical depression via smartphone sensing data In: *Proceedings of Wireless Health*; 2016 10 25; Bethesda, USA.
35. Yue C, Ware S, Morillo R, Lu J, Shang C, Bi J, et al. Fusing Location Data for Depression Prediction In: *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*; 2017 Aug 4–8; San Francisco, USA. New York (US): IEEE; 2018.
36. Lu J, Shang C, Yue C, Morillo R, Ware S, Kamath J, et al. Joint Modeling of Heterogeneous Sensing Data for Depression Assessment via Multi-task Learning. *Proc ACM Interact Mobile Wearable Ubiquitous Technol*. 2018;21 10.1145/3191753

37. Ware S, Yue C, Morillo R, Lu J, Shang C, Bi J, et al. Large-scale Automatic Depression Screening Using Meta-data from WiFi Infrastructure. *Proc ACM Interact Mobile Wearable Ubiquitous Technol.* 2018;2:195 10.1145/3287073
38. Ware S, Yue C, Morillo R, Lu J, Shang C, Bi J, et al. Predicting Depressive Symptoms Using Smartphone Data. *Smart Health.* 2020;15:100093.
39. Bardram JE, Frost M, Szanto K, Marcu G. The monarca self-assessment system: a persuasive personal monitoring system for bipolar patients. In: *Proceedings ACM SIGHIT International Health Informatics Symposium; 2012 Jan 28–30; Miami, USA. New York (USA): IEEE; 2012 p. 21–30.*
40. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram J, Kessing LV. Daily electronic monitoring of subjective and objective measures of illness activity in bipolar disorder using smartphones—the MONARCA II trial protocol: a randomized controlled single-blind parallel-group trial. *BMC Psychiatry* 2014;14:309. [PubMed: 25420431]
41. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. *J Med Internet Res.* 2011;13(3):e55. [PubMed: 21840837]
42. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: A pilot trial in the wild. *JMIR Mhealth Uhealth.* 2016;4(3):e111. [PubMed: 27655245]
43. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry.* 2003;54(5):573–83. [PubMed: 12946886]
44. Chen K, Xu T, Bi J. Latent sparse modeling of longitudinal multi-dimensional data. In: *Proceedings of AAAI Conference on Artificial Intelligence; 2018 Feb 2–7; New Orleans, USA.*
45. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Proceedings of the Neural Information Processing Systems Conference; 2014 Dec 8–13; Montréal Canada. San Diego (USA): NIPS Proceedings; 2014 p. 2672–80.*



**Figure 1.** DepWatch: high-level approach. The ground truth includes self-reported QIDS (Quick Inventory of Depressive Symptomatology) scores, and Monthly clinician assessment (including review of weekly QIDS scores and participant interview).



The figure consists of two side-by-side screenshots of a mobile application interface. The left screenshot shows two daily questionnaires. The first is titled "How would you rate your mood today?" and features a horizontal scale from 1 to 5 with a red dot at 3. Below it, the text reads "1: Excellent, 5: very bad". The second is titled "How would you rate your anxiety level today?" and also features a horizontal scale from 1 to 5 with a red dot at 3. Below it, the text reads "1: No Anxiety, 5: Severe Anxiety". A blue "SUBMIT" button is located below the second questionnaire. The right screenshot shows three weekly questionnaires. The first is titled "1. Overall in the past week, how would you rate the side-effects of the new depression medication?" and features a horizontal scale from "Not at All" to "Severe Side Effect" with a red dot at 3. The second is titled "2. Overall in the past week, how tolerable was the new depression medication?" and features a horizontal scale from "Not at All" to "Severe Side Effect" with a red dot at 3. The third is titled "3. In the past week, did you take the depression medication/s as prescribed by your provider?" and features five radio button options: "Always (100% of the time)", "Mostly (At least 75% of the time)", "Often (At least 50% of the time but less than 75% of the time)", "Sometimes (At least 25% of the time but less than 50% of the time)", and "Very little or Never (0 to 25% of the time)". A blue "SUBMIT" button is located below the third questionnaire.

**Figure 2.** Two daily questionnaires on mood and anxiety (**left**) and three weekly questionnaires on safety, tolerability, medication adherence (**right**).

**Table 1.**

Timeline of the project.

Tasks	Year 1		Year 2		Year 3		Year 4	
	1-6 months	7-12 months	1-6 months	7-12 months	1-6 months	7-12 months	1-6 months	7-12 months
Develop data collection system								
Phase I study recruitment								
Develop and evaluate machine learning models								
Develop web portal								
Phase II study recruitment								
Summarize lessons learned								

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript