

Detecting Depression of Cancer Patients with Daily Mental Health Logs from Mobile Applications

Completed Research Paper

Junetae Kim

KAIST

85 Hoegiro, Dongdaemun-gu, Seoul Korea
kaistar@business.kaist.ac.kr

Sanghee Lim

The Johns Hopkins Carey Business School
100 International Drive, Baltimore, MD
21202, USA
lim.sanghee@jhu.edu

Byungtae Lee

KAIST

85 Hoegiro, Dongdaemun-gu, Seoul Korea
btle@kaist.ac.kr

Jong Won Lee

Breast & Endocrine Division,
Department of Surgery
Asan Medical Center
88 Olympic-ro 43-gil, Songpa-gu, Seoul
Korea
jjjongwr@hanmail.net

Soo-Yong Shin

Department of Biomedical Informatics,
Asan Medical Center
88 Olympic-ro 43-gil, Songpa-gu, Seoul
Korea
sooyong.shin@ams.seoul.kr

Abstract

Mobile mental health trackers, the mobile applications that gather self-reported mental logs from users, have gained recent attention from clinicians as a tool for detecting patients' depression. However, critics have raised questions about the validity of the data collected from mental health trackers, which ask only a few simple questions using the face emoticon scale. Our research is the first study to address this issue, and we provide theoretical discussion that leads to the following hypotheses: (1) simpler but larger datasets collected daily from mobile mental health trackers can serve as good indicators to detect patients' depression, and (2) higher adherence to mobile mental health trackers enhances screening accuracy. We tested our hypotheses using the dataset of 5,792 sets of daily mental health logs collected from 78 breast cancer patients. Our random logistic panel regression and ROC analysis results, as well as k-means clustering analysis, provide strong supports for both hypotheses.

Keywords: Mobile health, Mental health trackers, PRO (Patient Reported Outcomes), Depression, ROC (Receiver Operating Characteristic), K-means clustering, Facial emoticon scales, Patient adherence

Introduction

Recently, self-report-based mobile health applications, which are designed to collect daily health logs of patients, have become actively used in many large hospitals (Judson et al. 2013; Min et al. 2014). In particular, in chronic and severe disease treatment settings, mobile mental health trackers have gained special attention as a tool for routine psychological distress screening (Donker et al. 2013; Harrison et al. 2011). Depression is a common symptom in patients in these settings, and early detection is imperative to prevent its adverse effects on patients' treatment processes and health outcomes (Katon et al. 2005; Pozuelo et al. 2009). However, several factors deter early detection, such as lack of feasible screening instruments in routine settings and time constraints of both patients and clinicians (Gjerdingen and Yawn 2007; Katon and Ludman 2003). The use of mobile mental health trackers is expected to help overcome those barriers and foster communication between patients and clinicians about patients' mental status without time and space restrictions (Chatterjee et al. 2009), providing richer datasets that are otherwise not easily available in traditional treatment settings. The recent advancement of EMR (electronic medical record) systems, which have the capacity to efficiently integrate data from different sources, has further heightened the hope that clinicians will be able to detect patients' mental status in real time based on daily logs reported via mental health trackers (Chatterjee et al. 2009). However, it is still unclear whether daily mental health log data, which use simpler instruments and are self-reported by patients through mobile mental health trackers, can be used to detect patients' mental status for clinical purposes. Moreover, critics raise questions about the validity of the use of the face emoticon scale used in many mobile mental health trackers for depression screening. In addition, from a practical perspective, the expansive amount of daily data collected from patients through mobile mental health trackers can be a burden, especially for a large health care system. Therefore, it is crucial to develop a framework to distinguish useful data from data that may only increase noise, bias, and variability, the common pitfalls of mobile data (Chen et al. 2012).

Our study aims to provide a theoretical framework and empirical evidence of the efficacy of the use of daily mental health logs collected through mobile mental health trackers to screen patients' mental status. We suggest that simpler but larger datasets collected daily from mobile mental health trackers can serve as good indicators to detect patients' depression. Also, based on the adherence literature in medical research, we suggest that the criteria of activeness, timeliness, and persistence in using mobile mental health trackers are useful to categorize patients into groups that show different levels of screening accuracy. We tested our hypotheses using the data collected from breast cancer patients who received treatment in the largest hospital in South Korea. Using the unique and novel dataset of 5,792 daily mental health logs of 78 breast cancer patients gathered via a mobile mental health tracker called "Pit-a-Pat" during a 48-week span, we developed a model that identifies patients' depression in three areas: sleep quality, mood, and anxiety levels. The results of a random logistic panel regression provide evidence that daily mental health logs serve as a reliable tool for screening patients for depression. Also, by employing a *k*-means clustering algorithm, we categorized the patients into two groups based on their adherence level, which is determined by their activeness, timeliness, and persistence in using the mobile mental health tracker. The results show that the accuracy is significantly higher for patients in the high-adherence group than those in the low-adherence group.

Our research has several strengths that provide important implications for both academic researchers and health care practitioners, in addition to the fact that our study is just one of a few studies that examine the mobile health trackers used by patients for clinical purposes. First, our study is the first to provide empirical evidence that daily mental health logs collected through mobile mental health trackers can serve as reliable indicators for detecting a patient's mental distress. Although plenty of studies have discussed the role of mobile health trackers in supporting clinicians' decision making, most of them focus on managerial suggestions and technical guidelines without sufficient empirical evidence (Albrecht 2013; Junglas et al. 2009). Empirical studies focus on evaluation of the feasibility of data collection without assessment of the usability of the data for clinical purposes (Harrison et al. 2011; Min et al. 2014; Reid et al. 2009). Our unique panel dataset of daily mental health logs combined with the screening results of traditional methods allows us to examine the validity of the mobile data for depression screening, while controlling for unobserved heterogeneity of patients or temporary external events that may cause estimation bias.

Second, our study is the first to empirically show the impact of patient adherence on health care quality in the mobile health trackers context. Although several studies mention the importance of patient adherence to reporting health logs via mobile health trackers (Katzan et al. 2011; Locklear et al. 2014; Snyder et al. 2009), these studies focus on strategies to promote patient adherence to applications and system designs for data collection on mobile devices. Our study is unique in the sense that we empirically show the positive effect of patient adherence to self-reporting via mobile mental health trackers on the screening accuracy of patients' mental status, extending prior studies that show the positive effect of patient adherence on health care outcomes in traditional healthcare settings (DiMatteo et al. 2002; Williams et al. 1998, 2002). Our empirical evidence will alleviate patients' burden in keeping daily logs by helping them understand the benefits of reporting daily via mobile mental health trackers (Locklear et al. 2014). Also, it may further motivate patients to adhere to the applications.

Third, we provide a theoretical discussion on the framework for the potential of mobile mental health trackers as effective depression screening tools and a methodological approach for restructuring the format of daily mental health logs to identify depression. Unlike traditional paper-based screening tools such as PHQ-9 (Patient Health Questionnaire-9), mobile mental health trackers gather data based on face emoticon scales, and both clinicians and users have raised concerns about the validity of this data, and, by extension, the validity of mobile mental health trackers as depression screening tools. Based on the psychological explanation about the positive effects of a short retention interval for mental status (Ayers and Reder 1998; Johnson et al. 1993; Odinet and Wolters 2006; Windschitl 1996) and studies confirming the accuracy of using face scales that express human emotions (Ekman et al. 1983; Goldman and Sripada 2005) in depression screening, we address how brief questions on face emoticon scales serve as similar results of depression screening with traditional tools. Furthermore, based on clinical guidelines for identifying depression (American Psychiatric Association 2013; Kroenke and Spitzer 2002), we propose a methodological approach to identify depression determined in a biweekly period with mental health logs gathered daily.

This paper is organized as follows. First, we provide the background of mobile mental health trackers and medical adherence. Next, we develop our hypotheses based on a review of relevant literature, followed by a description of the empirical design of this study. The results and discussion are provided.

Background

Mental Health Trackers for Depression in Oncologic Treatment

Although depression has adverse effects on the decision making process of patients, their treatment effectiveness, recovery, and their mortality risks (Stommel et al. 2002; Trask et al. 2001), mental distress, when present, is detected far less than 30% of the time in cancer patients, due to time constraints of both patients and clinicians and the lack of self-assurance in receiving depression screening tests (Gessler et al. 2008; Gil et al. 2005). To resolve these problems, brief screening tools consisting of just one or two self-report questionnaires, such as the Distress Thermometer (DT) and the Patient Health Questionnaire-2 (PHQ-2), are used (Gil et al. 2005; Mitchell and Coyne 2007). However, these screening methods are still problematic when dealing with those patients who rarely visit a doctor. To alleviate this issue, doctors recommend that such patients continue tracking Patient Reported Outcome (PRO) on paper as a form of mental status diary (Harrison et al. 2011; Turnbull Macdonald et al. 2012). However, due to the inconvenience of keeping daily logs on paper, the usage of such diaries is low (Harrison et al. 2011; Stone et al. 2002).

With the rapid surge in the use of mobile devices, health providers wish to take advantage of mobile technologies by embedding instruments that can collect mental PRO via mobile applications. Despite the potential benefits of mental health trackers in the oncologic treatment setting, prior studies have focused on evaluating the feasibility of data collection and overall response rates (Harrison et al. 2011; Min et al. 2014; Reid et al. 2009), without research on evaluating the validity of the data for depression screening. With this research, our intent is to close this gap.

Medication Adherence

In the medical research, adherence is defined as a patient's autonomous involvement in treatment plans¹ (Ho et al. 2009; Sandman et al. 2012) and is closely related to the concepts of "compliance" (Sandman et al. 2012).² Traditional medical research broadly studies the effects of adherence on health outcomes and prescribes strategies for improving adherence in the treatment process. For the effect on health outcomes, a vast amount of research has shown that non adherence to medication results in worse treatment outcome and health status, an increased rate of hospitalization, and higher health management costs (DiMatteo et al. 2002; Ruddy et al. 2009; Sokol et al. 2005; Waterhouse et al. 1993). Several studies have also discussed strategies to promote adherence, such as increasing shared decision making between doctors and patients and doctor-patient communication (Robinson et al. 2008; Street et al. 2009). Despite the importance of adherence, to the best of our knowledge, there has been no discussion on the effects of adherence to self-reported PRO via mobile health trackers in helping clinicians to better detect patients' mental status. We extend prior literature by considering adherence to self-report measures as the key factor determining the quality of PRO's.

Theoretical Framework and Hypothesis Development

Traditional Screening Tools vs. Mobile Mental Health Trackers

Mobile mental health trackers exhibit several distinctive characteristics compared with traditional screening tools for mental health assessment. Among them, two factors influence the accuracy of the screening results for patients' mental status: (1) the depth and breadth of the survey instrument and (2) the frequency of data collection. In short, mental health trackers collect data in much simpler forms but with higher frequency (i.e., daily) than traditional screening tools (Figure 1). We posit that a simpler form of survey may reduce the accuracy of the test results, but the higher frequency of data collection reduces potential measurement errors, offsetting the total impact.

While traditional tools such as PHQ-9 and BDI (Beck Depression Inventory) gather diverse information related to depressive symptoms on a long term basis (e.g, biweekly) (Beck and Alford 2009; Kroenke and Spitzer 2002), mobile mental health trackers ask patients to provide only a few items related to depressive symptoms on a short term basis (e.g., daily). Prior studies have shown that the accuracy of memory substantially decreases as the length of the retention period increases (Ayers and Reder 1998; Johnson et al. 1993; Odinet and Wolters 2006; Windschitl 1996). The memory retention issue is particularly critical for cancer patients, because their mental statuses are often unstable due to the side effects of cancer treatment (Badger et al. 2001; Vahdaninia et al. 2010). Thus, we expect that the use of mobile mental health applications can reduce memory errors. We also speculate that the use of a face emoticon scale has little significant harmful effect on the validity of the data and may in fact have a positive effect. A face emoticon scale is widely used in measuring pain and mental distress in mobile mental health trackers. Even though there has not been research directly examining the effectiveness of a face emoticon scale for detecting mental status in a mobile context, prior studies in psychology suggest that a face emoticon scale can be a good indicator of a patient's mental status, because a face emoticon scale demands less cognitive effort and less of a burden in interpreting the items (Bieri et al. 1990; McKinley et al. 2003). Goldman and Sripada (2005) have shown that people infer another person's mental status through his or her facial expression. Ekman et al. (1983) show through experiments that emotion can be reliably detected by facial muscle patterns. Also, a face emoticon scale can make the survey participation more enjoyable (Derham 2011). On the other hand, readability of a text-based rating scale is significantly reduced on mobile devices. Peytchev and Hill (2009) examine the survey design for mobile environments through experiments and show that heavy text-based information beyond the visible display bothers participants, causing some of them to ignore the text-based information. Therefore, the use of a face emoticon scale fitted on a small display size may even facilitate user participation, potentially making the data more useful. In conclusion,

¹ More specifically, the term "adherence" characterizes "patients as independent, intelligent, and autonomous people who take more active and voluntary roles in defining and pursuing goals for their medical treatment" (Lutfey and Wishner 1999; Sandman et al. 2012).

² The term "compliance" is defined as "the extent to which a person's behavior coincides with medical or health advice,"

we posit that a shorter survey instrument with a face emoticon scale is unlikely to diminish the accuracy of screening results of a mobile mental health tracker compared with traditional paper-based screening tools.

Based on these arguments, we hypothesize that the larger dataset collected through mobile mental health trackers capturing the measures of a few appropriate distress symptoms at a much shorter time interval (i.e., daily) will serve as a good indicator for patients' mental health status.

H1: Daily mental health logs reported via mobile mental health trackers provide screening results for depression consistent with those determined by using traditional instruments.

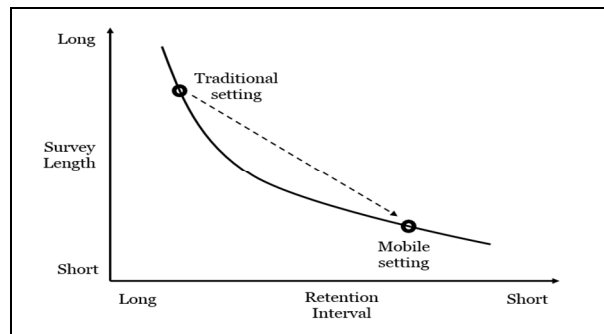


Figure 1. Difference in Data Collection between Traditional and Mobile Tools

Adherence to Self-Reporting and Its Effects on Screening Accuracy

Prior studies in medical research provide theoretical frameworks and empirical evidence that shows an association between a higher level of adherence to treatment plans and better health outcomes in the clinical setting. We posit that the positive effect of adherence on health care outcome can be extended to the context of mobile mental health trackers and suggest that higher adherence to mobile mental health tracking enhances screening accuracy for patients' depression.

The accuracy of a statistical prediction model is influenced by both quantity and quality of data (Guisasola et al. 2006), and we argue that a patient's adherence to self-reporting can influence both dimensions. Sandman et al. (2012) suggests that a patient is more likely to adhere to a treatment when the patient has chosen that particular treatment through their own preference. Thus adherent patients tend to make additional efforts to successfully accomplish the suggested treatment plan. Extending this argument, in our context, we argue that adherent patients are likely to be the ones most interested in managing their mental distress and are also willing to actively participate in this additional intervention. As a result, we expect that adherent patients not only report more logs in a larger quantity (e.g. reporting logs more frequently and persistently) but also report logs of higher quality, because these patients report the logs with more meticulous care.

We conceptualize adherence to mobile mental health tracking in three dimensions based on our review of prior literature: (1) activeness, (2) timeliness, and (3) persistence. First, activeness has been considered to be a primary indicator of a patient's adherence level in medical research and refers to the degree of the patient's activeness in adhering to guidelines (Lutfey and Wishner 1999; Sandman et al. 2012), often operationalized as the count of incidents of a patient's active participation (Cramer et al. 1995; Kardas 2005; Kronish et al. 2010; Paes et al. 1997). Second, timeliness captures a patient's behavior in reporting daily mental health logs without delay. The World Health Organization (WHO) has introduced on-time appointment-keeping as an index for measuring drug resistance (World Health Organization 2010). Prior studies have used timeliness to measure patients' attitude toward medication (Bastard et al. 2012; Blacher et al. 2010). Most daily logs gathered via mobile mental health trackers, including the applications used in this study, allow users to submit logs for the past few days. However, for cancer patients, delayed reports may contain biases, because they often experience high variations in their mental statuses due to the side effects of chemotherapy or medications (Badger et al. 2001; Vahdaninia et al. 2010). Therefore, reporting each set of mental health logs on the day of the report is highly recommended. Last, medication persistence is defined as continuous involvement with clinical treatment during the prescribed period

(Cramer et al. 2008). It is often characterized as a patient’s long-term attitude toward adhering to clinical guideline (Cramer et al. 2008; Lee et al. 2006) and is an important dimension of adherence, as routine depression screening is recommended throughout the whole cancer treatment period (Hopko et al. 2008; Pasquini and Biondi 2007; Tu et al. 2014). We consider a patient’s adherence to mobile mental health trackers as a composite construct of these three factors—activeness, timeliness, and persistency. They address different aspects of adherence, and there is no theoretical or empirical evidence that prioritizes the importance among the three. Moreover, the way patients adhere to mobile mental health trackers can vary across patients depending on their personalities. Some patients may prefer achieving short-term goals, while some patients are more enthusiastic about achieving long-term goals (Duckworth et al. 2007). In this regard, we still need to consider patients who submit only a few logs (low activeness) or submit the logs with delay (low timeliness) as having high adherence, if they are committed to using the applications during the entire course of the treatment (high persistence). Likewise, among patients who report daily logs frequently (high activeness) in over the long term (high persistence), we still need to distinguish patients who keep the logs on each day (high timeliness) from the ones who report the logs of past days on one day (low timeliness).

Their sincere attitude toward submitting their mental health logs through mobile mental health trackers is expected to have a positive influence on the intended health care outcome, which is the accuracy of the depression screening. Thus, we hypothesize that

H2: Depression screening accuracy is higher for patients who adhere to self-report measures, showing the active, timely, and persistent usage of mobile mental health trackers.

Data and Methodology

Data

In early 2013, the largest hospital in South Korea developed a mobile application called “Pit-a-Pat” aiming to promote interaction between health providers and breast cancer patients and collect PROs of those patients, such as mental health logs, drug history, and side effect of medication. The hospital started providing the application to female breast cancer patients who consented to use the application in April 2013. Among the PROs gathered by the application, we focus on three daily mental health logs: (1) anxiety, (2) mood, and (3) sleep satisfaction level, which have been demonstrated as symptoms or factors of depression in past literature (American Psychiatric Association 2013; Mayers et al. 2009; Seligman et al. 2001). Figure 2 displays screenshots from the application. Questionnaires on anxiety and sleep satisfaction are displayed as a visual thermometer, and a questionnaire for mood is displayed using a face emoticon scale. Sleep dissatisfaction³ is measured on a scale of 0 (very good) to 10 (very bad), mood level is recorded on a scale of 0 (none) to 7 (very severe), and anxiety level is measured on a scale of 0 (none) to 10 (very severe).

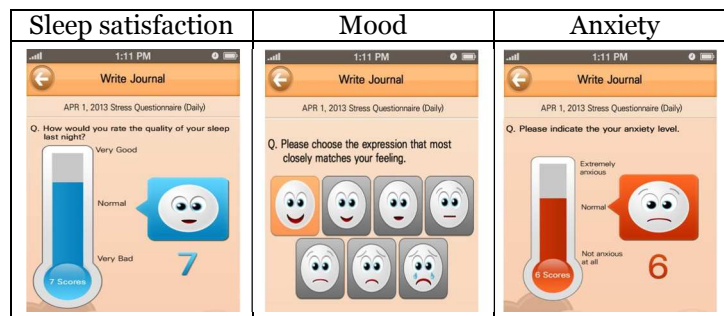


Figure 2. Snapshots of Three Self-Reporting Items

³ The application collects the sleep satisfaction level reported on a scale of 0 (very bad) to 10 (very good), but we reverse the scale to make it consistent with other measures (mood and anxiety) that take higher values as severity of depression increases.

Patient mental status is determined by results from PHQ-9, which is the most widely used depression screening tool in the primary care setting (Baldacci et al. 2013; Kravitz et al. 2013; Kroenke and Spitzer 2002; Lazenby et al. 2014). PHQ-9 consists of nine items, and each item is scored 0 to 3. The total score calculated by aggregating the scores of nine items has a value of 0 to 27, and represents severity of depression. The questionnaire items are provided in Appendix 1. PHQ-9 tests were administered on a biweekly basis via the Pit-a-Pat application.

Variables

Dependent Variable: Patient's Mental Status (Depressed)

We constructed the outcome variable, which indicates whether a patient is depressed in a two-week period, based on the PHQ-9 test results. The outcome variable, *Depressed*, takes the value of 0 if the PHQ-9 score is below 5 and one if the PHQ-9 score is greater than or equal to 5.

Severity of depressive symptoms on PHQ-9 is measured by the number of days that patients have symptoms related to depression over the two-week period (Kroenke and Spitzer 2002). Specifically, patients in the study score each severity of nine distress symptoms as 0 for "not at all," 1 as "two to six days," 2 as "seven to 13 days," and 3 as "every day," respectively. The scores of nine items are aggregated to determine the severity of depression, providing a total score of 0 to 27 (Kroenke and Spitzer 2002). As a treatment action guideline, five levels of depression severity are classified as "None" if scored 1 to 4, "Mild" if scored 5 to 9, "Moderate" if scored 10 to 14, "Moderately Severe" if scored 15 to 19, and "Severe" if scored 20 to 27 (Kroenke and Spitzer 2002). Along the 27-point scale, we used a conservative cutoff of 5 to determine depression, the criterion that has been widely used in prior research (Baldacci et al. 2013; Kravitz et al. 2013; Lazenby et al. 2014; McLennon et al. 2014). This conservative approach reduces the possibility that the cancer patients who have depressive symptoms are classified as normal. Several studies have reported that depression severity tends to be underestimated in the cancer treatment setting (Fann et al. 2008; Hardman et al. 1989; Hegel et al. 2006), despite the high cost of failing to detect depression due to its negative impact on health outcomes (Stommel et al. 2002; Trask et al. 2001). For this reason, in the depression treatment setting, researchers put more emphasis on improving a true-positive rate rather than a true-negative rate (Kroenke and Spitzer 2002), because it is far more important to correctly identify depressed people rather than correctly identify normal people. Therefore, we used the conservative cutoff value of 5.

Key Indicator Variable: Daily Mental Health Logs (Anxiety, Mood, Sleep)

As described earlier, mental logs were collected daily, while PHQ-9 tests were conducted biweekly. Due to the difference in the level of observations between mental logs (i.e. daily) and PHQ-9 results (i.e. biweekly), we reconstructed daily mental logs into a biweekly format. Three different approaches for aggregation are considered and we refer to them as the (1) average, (2) frequency, and (3) ratio approach. First, the average approach is used to measure the severity of depression during a two-week period by calculating the average of each type of mental logs of a patient during the period. However, because an average tends to be sensitive to outliers, practical guidelines suggest measuring the severity of depression by counting the number of days that people have symptoms related to the depression during specified periods (Association American Psychiatric 2013; Kroenke and Spitzer 2002). For this reason, for the latter two approaches, the frequency and the ratio approach respectively, we first determined the days with depressive symptoms by assigning a score of 1 to the days when the reported scores were above a certain cutoff value. For example, if a score of sleep quality on a day was higher than a cutoff value, say 7, we considered the patient to be depressed on the day and assign the value of 1 to the day for the patient. The optimal cutoff values for each type of mental health logs were determined by employing Receiver Operating Characteristic (ROC) analysis; the detailed procedures are provided in Appendix 2. The frequency approach counts the number of days with depressed symptom during a two-week period. The ratio approach calculates the ratio of the number of depressed days to the total number of days that the logs are reported during a two-week period.

Each approach includes pros and cons. As per discussion, the average approach was susceptible to outliers, and failed to account for mental states on the days when a patient did not report logs. The frequency

approach may underestimate depressive severity of patients who seldom report daily mental health logs, because it fails to account for omitted logs as well. On the other hand, the ratio approach may overestimate depressive severity when numerous omitted mental health logs are present. We consider that the ratio approach is more appropriate for our context, because correctly identifying the depressed state is preferred to correctly identifying the normal state (Kroenke and Spitzer 2002). Thus, we use the ratio approach for our main analysis, but we also display the results of the other approaches as well.

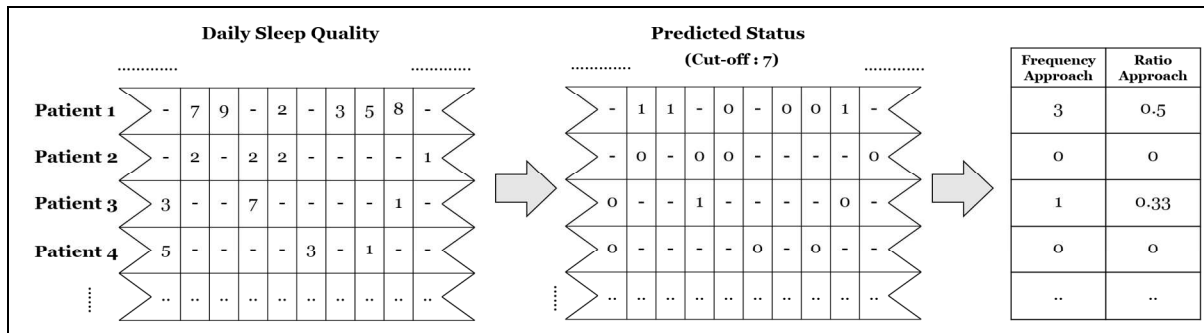


Figure 3. Illustration of Data Conversion from Daily Mental Health Logs to Biweekly Indicators

Classification of Patients According to Adherence Level

We classified patients into a higher adherence group and a lower adherence group based on three factors: activeness, timeliness, and persistence. Activeness is operationalized as the total number of days when daily mental health logs are reported. For timeliness, we measure the total number of days when the logs were reported on the same day of the report. Persistence is measured with two variables: (1) the number of biweekly periods between the first and last days with reported daily logs (i.e., total duration) and (2) the total number of biweekly periods with reported logs. The total duration is an important dimension of persistence, because it captures a discontinuity effect of the patients who stop using the application after a few weeks. However, there can be the case where a patient reports only two logs, one very early and the other later in the study period. Therefore, we also consider the number of biweekly periods with reported logs. It is still different from activeness, because this measure captures the low adherence of patients who reported logs very actively only during the first few weeks and then seldom used the application.

	Variable	Description
Depression Screening	Depressed	0 if normal (PHQ-9 score < 5) and 1 if depressed (PHQ-9 score ≥ 5)
	Sleep	The ratio of the number of depressed days determined by sleep log to the number of sleep logs reported during a period
	Mood	The ratio of the number of depressed days determined by mood log to the number of sleep logs reported during a period
	Anxiety	The ratio of the number of depressed days determined by anxiety log to the number of sleep logs reported during a period
Adherence Classification	Activeness	The total number of days when daily mental health logs are reported during the study period
	Timeliness	The total number of days when a patient reported mental health logs on the day of the report
	Persistence(1)	The total number of biweekly periods between the first and last observations
	Persistence(2)	The total number of biweekly periods with reported logs

Table 1. Description of Measurements

Model Specification and Clustering

Our model is to identify patients' mental status with three types of mental health logs:

$$\text{Depressed}_{i,t} = \text{Sleep}_{i,t} + \text{Mood}_{i,t} + \text{Anxiety}_{i,t} + \varepsilon_{i,t}$$

Subscripted i and t indicate each patient and each biweekly period, respectively. The dependent variable, *Depressed*, takes a binary value (0=normal, 1=depressed). Because our primary interest is to assess the extent to which daily mental health logs can identify patients' depression, we do not include control variables in our main model. However, we conduct a robustness check with the model that includes demographic information (cohabitation, education level, marital status, divorce status, age, the number of children, job status) to see if any important information is omitted.

The model parameters are estimated using a logistic random-effect regression. We use a logistic regression model because the dependent variable is a binary variable. For panel analysis, we employ a random-effect model instead of a fixed-effect model, because of the superior estimation efficiency of a random-effect model. As we will describe in the next section, our dataset is a short panel, meaning that the number of patients is far greater than the number of time span of observations. Therefore, estimation efficiency can be an issue with a fixed-effect model, because the model should also estimate the parameters of the dummy variables of which number is the same as the number of patients in our sample. It significantly reduces the degree of freedom of the estimation and makes the estimation relatively inefficient. Moreover, our dataset is unbalanced, meaning that it contains some patients who reported a PHQ-9 test result only once. All these patients will be dropped from analysis if the model is estimated with a fixed-effect model. Therefore, a random-effect model is preferred in our situation. A potential issue of a random-effect model is that the estimation parameters can be biased, if the indicator variables (sleep, mood, anxiety) covariate with the error terms due to unobserved patient-specific factors. As mentioned above, we test the model with various patient-specific factors as controls, and the results show little evidence that three types of mental health logs are endogenous. Robust standard errors are reported for potential heteroskedasticity issues.

We evaluate the screening accuracy of our model by employing Receiver Operating Characteristic (ROC) analysis. ROC is a graphical plot, which is widely used to demonstrate the prediction accuracy of a classifier model. It plots the true positive rate (i.e. sensitivity) against the false positive rate (i.e. 1-specificity) at various threshold values. Therefore, unlike a confusion matrix, it does not require researchers to select an arbitrary cut-off value to calculate the true/false positive rate, while providing all necessary information. The area under the ROC curve, which is referred to as an Area Under Curve (AUC), is interpreted as the probability that a classification model ranks a positive case higher than a negative case. Therefore, a higher AUC implies a better prediction performance of a classification model.

To classify patients based on their adherence level, we use a k -means clustering algorithm. K -means clustering classifies subjects into homogeneous subgroups, where each observation belongs to the cluster with the nearest intra-cluster distance and with the largest inter-cluster distance. The mechanism is to minimize the intra-cluster sum of squares as partitioning the n data into k heterogeneous subsets (mathematically, $\min S = \sum_{i=1}^k \sum_{x \in s_i} |x_n - \mu_i|^2$ where x_n represents n^{th} dimensional vector and μ_i is the mean of points). The procedure starts by randomly assigning x_n to k sets. Next, the mean of the cluster is calculated. Those steps are repeated until assigning x_n to k sets does not make any change.

Results

Sample Description

A total of 85 breast cancer patients consented and participated in this study, and they submitted 5,817 daily mental health logs from early April 2013 to late March 2014. We excluded 25 logs reported by seven patients who did not take any PHQ-9 tests during the sample period. As a result, 5,792 daily mental health logs reported by 78 patients were included in the analysis. The daily mental health log data were restructured according to the procedure described in the earlier section to match biweekly PHQ-9 test results. The 78 patients in our sample reported a total of 497 PHQ-9 test results, which consists of 270 normal statuses and 227 depressed statuses. Our data constitute an unbalanced panel of 78 patients for 24 biweekly periods. On average, we have 6.4 observations per patient. Eleven patients are observed only

once, and we have 24 observations for one patient. Table 2 provides summary statistics and a correlation matrix of key variables. For the 78 patients, we also collected their demographic information, including age, cohabitating status, number of children, marital status, education level, and employment status. Also, we include baseline information for patients' depression status measured using the BDI.

	N	Mean	Std.d	Min	Max	Depressed	Sleep	Mood	Anxiety
Depressed	497	0.46	0.50	0.00	1.00	1.00			
Sleep	497	0.21	0.23	0.00	1.00	0.36***	1.00		
Mood	497	0.40	0.38	0.00	1.00	0.42***	0.38***	1.00	
Anxiety	497	0.30	0.32	0.00	1.00	0.40***	0.22***	0.47***	1.00

Table 2. Summary Statistics and Correlation Matrix of Key Variables (Ratio Approach)

Dependent variable: <i>MentalStatus</i> , which is 0 if normal (PHQ-9 score < 5) and 1 if depressed (PHQ-9 score ≥ 5)					
	Model I	Model II	Model III	Model IV	Model V
Mental logs					
Sleep	3.712*** (0.890)			2.722*** (0.767)	2.957*** (0.734)
Mood		2.973*** (0.607)		1.783*** (0.596)	1.009* (0.599)
Anxiety			2.680*** (0.598)	1.782*** (0.661)	2.076*** (0.637)
BDI					0.104*** (0.033)
Demographic Controls ¹	No	No	No	No	Yes
Constant	-0.964*** (0.354)	-1.258*** (0.339)	-0.969*** (0.317)	-1.965*** (0.372)	-4.169** (1.498)
Num. of Obs.	497	497	497	497	496
Num. of Patients	78	78	78	78	77
Log likelihood	-264.1454	-258.0548	-265.479	-245.952	-234.4229
Wald χ^2	25.87***	39.00***	26.99***	58.31***	74.66***

1. The demographic controls are cohabitation, education level, marital status, divorce status, age, number of children, and employment status.
 Note: The 497 observations are constructed from a total of 5,817 daily mental health logs reported via the mental health tracking application. Random effects on patients. Robust standard errors are in parentheses. * significant at < 10%; **significant at < 5%; ***significant at < 1%.

Table 3. The Results of Random Effect Logistic Panel Regression

Random Effect Logistic Panel Regression Results (Ratio Approach)

Table 3 provides the random-effect logistic panel regression results. First, we examine the statistical power of each type of mental health log (sleep, mood, anxiety). Models I to III are the models that identify patients' mental status using one type of mental health log. The Wald χ^2 test results support our finding that all models are statistically valid. The coefficient of the mental health log in each model is statistically significant at the 1% level, supporting our hypotheses that the variables constructed using daily mental health logs are statistically significant indicators for patients' mental status. Next, we examine the effects of all three types of mental health logs in a single model. In Model IV, all three types of mental health logs are statistically significant in detecting patients' mental status at the 1% significance level. This result indicates that each type of mental health log addresses different dimensions of patient mental status. For example, consider the case when two patients reported the same level of anxiety and mood condition but

differing sleep conditions. Our result suggests that the difference in their sleep condition is a significant factor that determines their mental status. Holding other variables fixed, a one-tenth unit (0.1) increase in *Sleep* (i.e., the increase in the ratio of depressed day to the total number of reported days in a given biweekly period by 0.1) is associated with a 31.3% increase in the odd of the patient being depressed since $\exp(0.272) = 1.313$. Similarly, all other things being equal, a one-tenth unit increase in *Mood* and *Anxiety* is associated with a 19% increase in the odd of the patient being depressed, respectively. The LR test comparing Model IV with Models I–III shows that the integrated approach (Model IV), which uses all three types of mental health logs, explains depression better than the single-indicator model at the 1% significance level. Last, in Model V, we include demographic information and the BDI at the start of using the application as controls. The results show that all demographic controls are statistically insignificant. Consistent with our expectation, the coefficient of BDI is statistically significant at the 1% significance level, suggesting that the initial depression level of a patient is statistically positively associated with the depression level in the study period. Even after controlling for various patient-specific factors, mental health log variables are still statistically significant at the 10% level.

ROC Analysis Results

Figure 4-(1) shows the ROC plots of the prediction results obtained from the logistic regression analysis in Table 3 and the corresponding AUC's. Each line shows the ROC of Models I–IV. The AUC's calculated from the ROC of integrated logs, a sleep log, a mood log, and an anxiety log are 0.8012, 0.7105, 0.7403, and 0.7340, respectively. This result is consistent with the LR test result, which shows a superior model fit of Model IV. The differences between the AUCs of Model IV and other models are statistically significant at the 1% level, suggesting that the screening accuracy of Model IV is significantly higher than the single-indicator models. In particular, the slope of ROC of Model IV is steeper than the others, indicating that Model IV can correctly identify depressed patients as depressed (i.e., a high true-positive rate) with a lower risk of identifying normal patients as depressed (i.e., a low false-positive rate) than the other models.

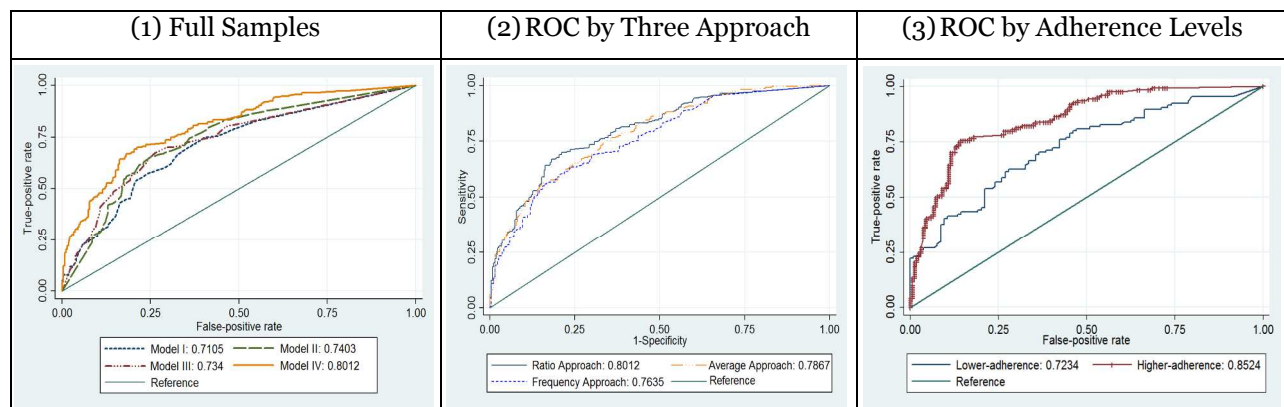


Figure 4. ROC Analysis

We conducted additional analyses to check the robustness of our results. First, we conducted the regression and ROC analysis using the alternative approaches for restructuring daily data using a biweekly format. Figure 4-(2) shows the results of the average approach and the frequency approach along with the results of the ratio approach, which are identical to the results of Model IV in Figure 4-(1). The results of the average and the frequency approach are generally consistent with those of the ratio approach. For the average approach, the coefficients of all three types of logs are statistically significant at the 5% significant levels, and the AUC of the ROC curves is 0.7867. For the frequency approach, the coefficients of *Sleep* and *Mood* are statistically significant, and the AUC of the ROC is 0.7634. The results of these alternative approaches suggest that the validity of the use of daily mental health log data is robust to the different approaches. The results also suggest that the prediction accuracy of the model with the ratio approach is the highest, supporting our selection of the ratio approach for our main analysis. The

AUC with the ratio approach is higher than the ones of the other approaches. Even though the difference between the ratio and average approach is not statistically different, it is statistically different between the ratio and the frequency approach.

Second, we tested the validity of our model by employing the fivefold cross-validation procedure to examine whether our models overfit the data. We (1) randomly partition the data into five subsets where the sample size is approximately 100, (2) run a random effect logistic regression using four of the subsets as a training set, (3) then calculate the predicted probability for the remaining subset as a test dataset, and (4) employ ROC analysis and calculate the AUC. Steps (1) through (4) are repeated five times by alternating training and test datasets. For the ratio approach, the resulting AUCs of the five subsets range from 0.7537 to 0.8568. The AUC of the aggregated results of the five subsets is 0.7937. For the average approach, the AUCs range from 0.7234 to 0.8488, and the AUC of the aggregated result is 0.7755. The resulting AUCs of the frequency approach range from 0.7177 to 0.8188, and the AUC of the aggregated result is 0.7550. The results suggest that the risk of overfitting is not high with our models, as the screening accuracy does not significantly decrease when the model is applied to the dataset not used for building a model.

Patient Adherence and Screening Accuracy

We analyzed the difference in depression screening accuracy between high and low adherence groups, which are classified by the *k*-means clustering algorithm.

	# of patients	# of obs.	Activeness	Timeliness	Duration	Persistence
Total	78	497	68.55	51.31	6.59	6.45
Lower adherence	58	208	37.66	29.81	3.79	3.67
Higher adherence	20	289	158.15	113.65	14.70	14.50
ANOVA Test* : F (1,76)	-	-	265.25***	176.29***	171.56***	181.08***

* Ho: Mean of lower adherence = Mean of higher adherence

Table 4. Mean of Key Variables by Groups Classified by K-means Clustering

Table 4 provides descriptive statistics of four variables for the two groups classified with the *k*-means clustering algorithm. “Higher adherence” refers to the group for which the mean values of the four variables are higher, and “lower adherence” refers to the other group. Among 78 patients, 58 patients are classified into the lower-adherence group, and 20 patients are classified into the higher-adherence group. Among the 497 observations in the biweekly panel dataset, there are 208 observations for the patients in the low-adherence group, and 289 observations for the patients in the high-adherence group. The ANOVA test results shows that the difference between the two groups are statistically significant at the 1% level.

Figure 4-(3) shows the ROC curves and the statistical test results that compare the screening accuracy according to two groups of adherence levels (high and low). With the ratio approach, AUCs of the higher-adherence group (0.8524) are significantly higher than the AUCs of the lower adherence group (0.7234) at the 1% significance level. These results show that the depression screening accuracy of patients who adhere to self-reporting on daily mental health logs is higher than the accuracy of patients who do not. The results with the average approach (the higher adherence: 0.8425, the lower adherence: 0.7016) and the frequency approach (the higher adherence: 0.8259, the lower adherence: 0.6664) are consistent with our findings.

One may suggest that there can be another factor that affects both depression and adherence, causing potential endogeneity issues. For example, what if patients with depression tend to adhere to self reporting?⁴ We examined such a possibility by analyzing whether other available variables such as a patient’s demographic information (i.e. age, cohabitating status, number of children, marital status, education level, and occupation status) and baseline mental information (BDI) were associated with their adherence level. We conducted t-test and Pearson’s χ^2 test, and the results show that the adherence level

⁴ We appreciate the anonymous reviewer for this point.

is not statistically associated with any of variables we tested. The random effect logit analysis which use the clustering result (1=the higher adherence, 0=the lower adherence) as a dependent variable also supports that no other factors except their marital status are related to an adherence level at the 10% significance level. When marital status is incorporated into the model, the model fails to converge. These results suggest that the risk of endogeneity due to other factors affecting depression and adherence is not high.

Comparisons	Layers	Adherence	Average Approach		Ratio Approach		Frequency Approach	
			AUC	Sig.	AUC	Sig.	AUC	Sig.
The Comparison of AUCs in Six Months and Without Late Starters	Six Months	Lower (n=161)	0.7239	0.033	0.728	0.016	0.6774	0.007
		Higher (n=336)	0.8205		0.8364		0.8088	
	Without Late Starters	Lower (n=171)	0.7015	0.001	0.7405	0.006	0.6796	0.002
		Higher (n=273)	0.8540		0.8599		0.8283	
The Comparison of AUCs by Only a Single Mental Health Log for the Study Period	Sleep	Lower (n=208)	0.5938	0.000	0.6102	0.001	0.5811	0.000
		Higher (n=289)	0.8160		0.7746		0.7639	
	Anxiety	Lower (n=208)	0.6922	0.048	0.6703	0.029	0.642	0.011
		Higher (n=289)	0.7816		0.7724		0.7629	
	Mood	Lower (n=208)	0.6802	0.000	0.6783	0.035	0.656	0.002
		Higher (n=289)	0.8353		0.7752		0.7985	
The Comparison of AUCs According to Three	Three Groups	Lower (n=113)	0.7134	0.020	0.6767	0.006	0.6003	0.002
		Middle (n=159)	0.7542		0.7893		0.7986	
		Higher (n=225)	0.8446		0.8512		0.8114	

Table 5. Robustness Test Results

We also conducted four additional tests to check the robustness of our finding that the screening accuracy is higher for patients with a higher level of adherence. First, we examined whether the difference in length of data collection periods by patients influences the results. Our main analysis considers the whole study period (48 weeks) to examine the effect of patient adherence. However, each patient started using the application at a different time during the study period, and the measure of persistence can be biased for patients who started using the application very early or very late. For example, persistence can be underestimated for patients who started using the application later in the study period, because they simply had fewer days to report their daily logs. Likewise, persistence can be overestimated for patients who started using the application earlier, because they spent more days with the application. Therefore, we evaluated whether our results are maintained if we consider only the log data collected during the first 24-week period for each patient. We re-ran the analysis, and the results are provided in Table 5. The result is consistent with our main result. Second, as an alternative way to address the concerns for the different time frames for patients, we repeated the analysis with the data subset by excluding patients who joined the study during the last 12 weeks, which is the average usage period of the patients in our sample. The result is provided in Table 5 and is consistent with our main results. Third, we examined whether the result holds when only one type of mental health log is used for the detection. The result in Table 5 shows that the effect of adherence on screening accuracy is maintained when only one type of mental health log is used for the detection. Fourth, we examined whether the results are maintained when patients are classified into three groups instead of two. This analysis is conducted to address the concern for outliers in each group (high and low), which may have driven the results. We re-clustered patients into three groups based on four variables. The ANOVA test results support that the differences between three groups are statistically significant at the 1% level. The results still support that the screening accuracy is higher for patients with a higher level of adherence still holds ($p < 0.05$) with the three-level classification (Table 5).

Discussion

Our study has several limitations that provide opportunities for future research. First, we determine patients' status based not on doctor's diagnosis results but on a screening instrument, PHQ-9 in this study. Therefore, our result should not be interpreted as the usefulness of mobile mental health trackers as a diagnostic tool for depression. Instead, we show that a mobile mental health tracker can be used as a screening tool to identify patients who have depressive symptoms. Therefore, for future research, it will be worthwhile to evaluate whether mobile mental health trackers can be used as a tool to diagnose patients' mental distress by comparing the results with doctors' clinical diagnoses. Second, although we conduct cross-validation tests and several robustness checks, our approach may still have a potential overfitting issue. We account for the number of depressive days that patients experienced during the specified period to construct explanatory variables, and this approach is similar to how a PHQ-9 test evaluates depressive severity. Therefore, although our study paves an important step in mobile mental health tracker research by proposing a new methodological approach to use daily mental health logs for depression screening, we believe a methodological improvement for alternative specifications for measurements and a detection model is an interesting area to investigate. Third, we consider three types of mental health logs collected from the application considered in this study. However, there is no standard way to collect mental health logs, and there has been no study on how depression questionnaires should be designed for the mobile environment. Some mobile mental health trackers ask only one brief, vague question, such as "how are you today?" Although short instruments are recommended to reduce burdens of reporting PRO (Locklear et al. 2014), our empirical results show that the depression screening accuracy is highest when all three types of logs are used rather than when only one type of log is used. Therefore, a natural extension of our study will be investigating the optimal design of mobile mental health trackers. Future study may also consider the effect of push alerts to promote adherence to self-reporting; mobile features like SMS messages have been shown to be effective in promoting adherence to medication and treatment plans (Armstrong et al. 2009; Free et al. 2013; Strandbygaard et al. 2010). Fourth, our study is conducted in a specific disease treatment setting in South Korea—depression screening for breast cancer patients. Therefore, our empirical evidence may not be generalized for other types of mental illness or for patients with different diseases, especially for patients with more severe disease, such as pancreatic and rectal cancers. Furthermore, South Korea has one of the highest percentages of smartphone users compared with other countries⁵. Accordingly, application development technology and data management skills there are considered to be of high quality. Therefore, the implications from our study may not be applied in an environment where complementary infrastructures are not adequately supported. In this regard, our study warrants further research on the assessment of the use of mobile mental health trackers in other settings. Last, many wearable devices are deployed by mobile healthcare applications. Such technology diffusion may not only further reduce patients' effort to report their health status but also improve the quality of signals by automating capturing of them. We plan to extend our study with wearables.

Despite these limitations, our study contributes to the academic research in several ways and provides important practical implications for health care providers. First, our study is the first attempt to identify mental distress based on daily mental health logs gathered through mobile mental health trackers. Although several mental health applications have been adopted as a means to overcome the limitations of traditional instruments for routinized depression screening (Min et al. 2014; Reid et al. 2009), there has been little discussion on the performance of mobile mental health trackers for clinical purposes. In particular, critics question the validity of relatively simple questionnaires and use of face emoticons by mobile mental health trackers. Our results show that depression screening using daily mental health logs via mobile mental health trackers can be a reasonable alternative to screen patients' depression, supporting the potential of mobile mental health trackers for early detection of depression. Second, we provide a new perspective on measuring adherence to self-reporting by using a multidimensional construct, consisting of activeness, timeliness, and persistence. Prior empirical studies on adherence to mobile PRO tend to focus only on activeness (e.g., the total number of logs), without considering that an overall adherence level can decrease as time goes by (Judson et al. 2013; Min et al. 2014). By incorporating the degree of a patient's autonomy to report on the right occasion (timeliness) during the whole treatment period (persistence), our framework enables us to capture the time effects in both the short-term and long-term horizons, an element that has been missing in prior research. Third, our

⁵ <http://think.withgoogle.com/mobileplanet/en/>

empirical evidence emphasizes the critical role of adherence to self-reporting, providing important lessons for both doctors and patients. Our results show that the depression screening accuracy of mental health logs is much higher for high-adherence groups (AUC: 0.8524) than for lower-adherence groups (AUC: 0.7234). Reporting daily mental health logs can be a significant burden for patients and can have an adverse effect on their mental status (Donaldson 2004; Snyder et al. 2009). These burdens can be reduced if patients recognize the clinical benefits of reporting their outcomes (PRO) (Locklear et al. 2014). Our empirical evidence can help patients understand the positive effect of adherence and provide motivation to adhere to self-reporting to improve the quality of treatment.

Because mental illness is a subjective disease and cannot be detected based on biological states such as blood pressure and body temperature, psychiatrists identify depressed patients based on patient testimony. Mental health trackers can be a good solution for connecting patients and doctors and promoting their communications. Our study lays the foundation for research on mental health trackers in the clinical setting. We provide empirical evidence of mental health trackers as a depression screening tool and highlight the importance of patients' active participation in mental treatment plans. We hope to see further research on mental health trackers, which are interesting and innovative approaches that will benefit both patients and doctors.

Appendices

A1. PHQ-9 Questionnaire Items (Kroenke and Spitzer 2002)

1. Little interest or pleasure in doing things; 2. Feeling down, depressed, or hopeless; 3. Trouble falling or staying asleep, or sleeping too much; 4. Feeling tired or having little energy; 5. Poor appetite or overeating; 6. Feeling bad about yourself, or that you are a failure or have let yourself or your family down; 7. Trouble concentrating on things, such as reading the newspaper or watching television; 8. Moving or speaking so slowly that other people could have noticed, or the opposite—being so fidgety or restless that you have been moving around a lot more than usual; 9. Thoughts that you would be better off dead or of hurting yourself in some way.

A2. Calculating Cutoff Values

To find the optimal cutoff value, we conduct ROC analysis. We calculate the optimal cutoff value of each mental health log by running a model in which a mental status is identified by only one type of mental health log with an arbitrary cutoff value and comparing AUC for all possible cutoff values. For example, to determine the cutoff for the *Sleep* variable, we (1) select an arbitrary cutoff value, c , (2) calculate the ratio or the frequency to get the *Sleep* variable as described in the measurement section (see Figure 3), (3) estimate the following model, $Depressed_{i,t} = Sleep_{i,t} + \epsilon_{i,t}$, and (4) get AUC from ROC analysis. As the sleep log data can take the value from 0 to 9, we obtain nine AUCs by repeating this process nine times. Among the nine candidate cutoff values from one to nine, we select the cutoff values that maximize AUC as the optimal cutoff values. For both the ratio and frequency approaches, the optimal cutoff values for sleep, anxiety, and mood are identified as seven, six, and four, respectively.

References

- Albrecht, U.-V. 2013. "Transparency of Health-Apps for Trust and Decision Making.," *Journal of Medical Internet Research* (15:12), p. e277 (doi: 10.2196/jmir.2981).
- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*, Washington, DC: American Psychiatric Association.
- Armstrong, A. W., Watson, A. J., Makredes, M., Frangos, J. E., Kimball, A. B., and Kvedar, J. C. 2009. "Text-Message Reminders to Improve Sunscreen Use: A Randomized, Controlled Trial Using Electronic Monitoring.," *Archives of Dermatology* (145:11), pp. 1230–6 (doi: 10.1001/archdermatol.2009.269).
- Ayers, M. S., and Reder, L. M. 1998. "A Theoretical Review of The Misinformation Effect: Predictions from an Activation-Based Memory Model," *Psychonomic Bulletin & Review* (5:1), pp. 1–21 (doi: 10.3758/BF03209454).
- Badger, T. A., Braden, C. J., and Mishel, M. H. 2001. "Depression Burden, Self-Help Interventions, and Side Effect Experience in Women Receiving Treatment for Breast Cancer.," *Oncology Nursing Forum* (Vol. 28).
- Baldacci, F., Vedovello, M., Ulivi, M., Vergallo, A., Poletti, M., Borelli, P., Cipriani, G., Nuti, A., and Bonuccelli, U. 2013. "Triggers in Allodynic and Non-Allodynic Migraineurs. A Clinic Setting Study.," *Headache* (53:1), pp. 152–60 (doi: 10.1111/head.12012).
- Bastard, M., Pinoges, L., Balkan, S., Szumilin, E., Ferreyra, C., and Pujades-Rodriguez, M. 2012. "Timeliness of Clinic Attendance is a Good Predictor of Virological Response and Resistance to Antiretroviral Drugs in HIV-Infected Patients.," *PloS One* (7:11), p. e49091 (doi: 10.1371/journal.pone.0049091).
- Bieri, D., Reeve, R. a., Champion, D. G., Addicoat, L., and Ziegler, J. B. 1990. "The Faces Pain Scale for the Self-Assessment of the Severity of Pain Experienced by Children: Development, Initial Validation, and Preliminary Investigation for Ratio Scale Properties," *Pain* (41:2), pp. 139–150 (doi: 10.1016/0304-3959(90)90018-9).
- Blacher, R. J., Muiruri, P., Njobvu, L., Mutsotso, W., Potter, D., Ong'ech, J., Mwai, P., Degroot, A., Zulu, I., Bolu, O., Stringer, J., Kiarie, J., and Weidle, P. J. 2010. "How Late Is Too Late? Timeliness to Scheduled Visits as an Antiretroviral Therapy Adherence Measure in Nairobi, Kenya and Lusaka, Zambia.," *AIDS Care* (22:11), pp. 1323–31 (doi: 10.1080/09540121003692235).
- Chatterjee, S., Chakraborty, S., Sarker, S., Sarker, S., and Lau, F. Y. 2009. "Examining the Success Factors for Mobile Work in Healthcare: A Deductive Study," *Decision Support Systems* (46:3), Elsevier B.V., pp. 620–633 (doi: 10.1016/j.dss.2008.11.003).
- Chen, C., Haddad, D., Selsky, J., Hoffman, J. E., Kravitz, R. L., Estrin, D. E., and Sim, I. 2012. "Making Sense of Mobile Health Data: An Open Architecture to Improve Individual- and Population-Level Health," *Journal of Medical Internet Research* (14:4), p. e112 (doi: 10.2196/jmir.2152).
- Cramer, J. a, Roy, A., Burrell, A., Fairchild, C. J., Fuldeore, M. J., Ollendorf, D. a, and Wong, P. K. 2008. "Medication Compliance and Persistence: Terminology and Definitions.," *Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research* (11:1), pp. 44–7 (doi: 10.1111/j.1524-4733.2007.00213.x).
- Cramer, J., Vachon, L., Desforges, C., and Sussman, N. M. 1995. "Dose Frequency and Dose Interval Compliance with Multiple Antiepileptic Medications during a Controlled Clinical Trial," *Epilepsia* (36:11), pp. 1111–1117 (doi: 10.1111/j.1528-1157.1995.tb00469.x).
- Derham, P. A. J. 2011. "Using Preferred, Understood or Effective Scales? How Scale Presentations Effect Online Survey Data Collection," *Australasian Journal of Market & Social Research* (19:2).
- DiMatteo, M. R., Giordani, P. J., Lepper, H. S., and Croghan, T. W. 2002. "Patient Adherence and Medical Treatment Outcomes: A Meta-Analysis.," *Medical Care* (40:9), pp. 794–811 (doi: 10.1097/01.MLR.0000024612.61915.2D).
- Donaldson, M. S. 2004. "Taking Stock of Health-Related Quality-of-Life Measurement in Oncology Practice in the United States.," *Journal of the National Cancer Institute. Monographs* (33), pp. 155–67 (doi: 10.1093/jncimonographs/lgh017).

- Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., and Christensen, H. 2013. "Smartphones for Smarter Delivery of Mental Health Programs: a Systematic Review.," *Journal of Medical Internet Research* (15:11), p. e247 (doi: 10.2196/jmir.2791).
- Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. 2007. "Grit: Perseverance and Passion for Long-Term Goals.," *Journal of Personality and Social Psychology* (92:6), pp. 1087–101 (doi: 10.1037/0022-3514.92.6.1087).
- Ekman, P., Levenson, R., and Friesen, W. 1983. "Autonomic Nervous System Activity Distinguishes among Emotions," *Science* (221:4616), pp. 1208–1210 (doi: 10.1126/science.6612338).
- Fann, J. R., Thomas-Rich, A. M., Katon, W. J., Cowley, D., Pepping, M., McGregor, B. a, and Gralow, J. 2008. "Major Depression after Breast Cancer: A Review of Epidemiology and Treatment.," *General Hospital Psychiatry* (30:2), pp. 112–26 (doi: 10.1016/j.genhosppsych.2007.10.008).
- Free, C., Phillips, G., Galli, L., Watson, L., Felix, L., Edwards, P., Patel, V., and Haines, A. 2013. "The Effectiveness of Mobile-Health Technology-Based Health Behaviour Change or Disease Management Interventions for Health Care Consumers: A Systematic Review.," *PLoS Medicine* (10:1), p. e1001362 (doi: 10.1371/journal.pmed.1001362).
- Gessler, S., Low, J., Daniells, E., Williams, R., Brough, V., Tookman, A., and Jones, L. 2008. "Screening for Distress in Cancer Patients: Is the Distress Thermometer a Valid Measure in the UK and Does It Measure Change over Time? A Prospective Validation Study," *Psycho-Oncology* (17:6), pp. 538–547 (doi: 10.1002/pon).
- Gil, F., Grassi, L., Travado, L., Tomamichel, M., and Gonzalez, J. R. 2005. "Use of Distress and Depression Thermometers to Measure Psychosocial Morbidity among Southern European Cancer Patients.," *Supportive Care in Cancer: Official Journal of the Multinational Association of Supportive Care in Cancer* (13:8), pp. 600–6 (doi: 10.1007/s00520-005-0780-0).
- Gjerdingen, D. K., and Yawn, B. P. 2007. "Postpartum Depression Screening: Importance, Methods, Barriers, and Recommendations for Practice.," *Journal of the American Board of Family Medicine : JABFM* (20:3), pp. 280–8 (doi: 10.3122/jabfm.2007.03.060171).
- Goldman, A. I., and Sripada, C. S. 2005. "Simulationist Models of Face-Based Emotion Recognition.," *Cognition* (94:3), pp. 193–213 (doi: 10.1016/j.cognition.2004.01.005).
- Guisasola, a., Baeza, J. a. A., Carrera, J., Sin, G., Vanrolleghem, P. a. A., and Lafuente, J. 2006. "The influence of experimental data quality and quantity on parameter estimation accuracy," *Education for Chemical Engineers* (1:1), pp. 139–145 (doi: 10.1205/ece06016).
- Hadidi, N., Treat-Jacobson, D. J., and Lindquist, R. 2009. "Poststroke Depression and Functional Outcome: A Critical Review of Literature," *Heart & Lung: The Journal of Acute and Critical Care* (38:2), pp. 151–162 (doi: 10.1016/j.hrtlng.2008.05.002).
- Hardman, A., Maguire, P., and Crowther, D. 1989. "The Recognition of Psychiatric Morbidity on a Medical Oncology Ward.," *Journal of Psychosomatic Research* (33:2), pp. 235–239 (doi: 10.1016/0022-3999(89)90051-2).
- Harrison, V., Proudfoot, J., Wee, P. P., Parker, G., Pavlovic, D. H., and Manicavasagar, V. 2011. "Mobile Mental Health: Review of the Emerging Field and Proof of Concept Study.," *Journal of Mental Health (Abingdon, England)* (20:6), pp. 509–24 (doi: 10.3109/09638237.2011.608746).
- Hegel, M. T., Moore, C. P., Collins, E. D., Kearing, S., Gillock, K. L., Riggs, R. L., Clay, K. F., and Ahles, T. a. 2006. "Distress, Psychiatric Syndromes, and Impairment of Function in Women with Newly Diagnosed Breast Cancer.," *Cancer* (107:12), pp. 2924–31 (doi: 10.1002/cncr.22335).
- Ho, P. M., Bryson, C. L., and Rumsfeld, J. S. 2009. "Medication Adherence: Its Importance in Cardiovascular Outcomes.," *Circulation* (119:23), pp. 3028–35 (doi: 10.1161/CIRCULATIONAHA.108.768986).
- Hopko, D. R., Bell, J. L., Armento, M. E. A., Robertson, S. M. C., Hunt, M. K., Wolf, N. J., and Mullane, C. 2008. "The Phenomenology and Screening of Clinical Depression in Cancer Patients.," *Journal of Psychosocial Oncology* (26:1), pp. 31–51 (doi: 10.1300/JO77v26n01_03).
- Jani, B. D., Purves, D., Barry, S., Cavanagh, J., McLean, G., and Mair, F. S. 2013. "Challenges and Implications of Routine Depression Screening for Depression in Chronic Disease and Multimorbidity: A Cross Sectional Study.," *PloS One* (8:9), p. e74610 (doi: 10.1371/journal.pone.0074610).
- Johnson, M. K., Hashtroudi, S., and Lindsay, D. S. 1993. "Source Monitoring.," *Psychological Bulletin* (114:1), pp. 3–28 (available at <http://www.ncbi.nlm.nih.gov/pubmed/8346328>).
- Judson, T. J., Bennett, A. V, Rogak, L. J., Sit, L., Barz, A., Kris, M. G., Hudis, C. a, Scher, H. I., Sabattini, P., Schrag, D., and Basch, E. 2013. "Feasibility of Long-Term Patient Self-Reporting of Toxicities

- from Home via the Internet during Routine Chemotherapy.,” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* (31:20), pp. 2580–5 (doi: 10.1200/JCO.2012.47.6804).
- Junglas, I., Abraham, C., and Ives, B. 2009. “Mobile Technology at the Frontlines of Patient Care: Understanding Fit and Human Drives in Utilization Decisions and Performance,” *Decision Support Systems* (46:3), pp. 634–647 (doi: 10.1016/j.dss.2008.11.012).
- Kardas, P. 2005. “The DIACOM Study (Effect of Dosing Frequency of Oral Antidiabetic Agents on the Compliance and Biochemical Control of Type 2 Diabetes).,” *Diabetes, Obesity & Metabolism* (7:6), pp. 722–728 (doi: 10.1111/j.1463-1326.2004.00462.x).
- Katon, W. J., and Ludman, E. J. 2003. “Improving Services for Women with Depression in Primary Care Settings,” *Psychology of Women Quarterly*, pp. 114–120 (doi: 10.1111/1471-6402.00091).
- Katon, W. J., Rutter, C., Simon, G., Lin, E. H. B., Ludman, E., Ciechanowski, P., Kinder, L., Young, B., and Von Korff, M. 2005. “The Association of Comorbid Depression With Mortality in Patients With Type 2 Diabetes,” *Diabetes Care* (28:11), pp. 2668–2672 (doi: 10.2337/diacare.28.11.2668).
- Katzan, I., Speck, M., Dopler, C., Urchek, J., Bielawski, K., Dunphy, C., Jehi, L., Bae, C., and Parchman, A. 2011. “The Knowledge Program: An Innovative, Comprehensive Electronic Data Capture System and Warehouse.,” *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium* (2011), pp. 683–92 (available at <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3243190&tool=pmcentrez&rendertype=abstract>).
- Kravitz, R. L., Franks, P., Feldman, M. D., Tancredi, D. J., Slee, C. a, Epstein, R. M., Duberstein, P. R., Bell, R. a, Jackson-Triche, M., Paterniti, D. a, Cipri, C., Iosif, A.-M., Olson, S., Kelly-Reif, S., Hudnut, A., Dvorak, S., Turner, C., and Jerant, A. 2013. “Patient Engagement Programs for Recognition and Initial Treatment of Depression in Primary Care: A Randomized Trial.,” *JAMA : the journal of the American Medical Association* (310:17), pp. 1818–28 (doi: 10.1001/jama.2013.280038).
- Kroenke, K., and Spitzer, R. R. L. 2002. “The PHQ-9: A New Depression Diagnostic and Severity Measure,” *Psychiatric Annals* (32:9), pp. 1–7.
- Kronish, I. M., Rieckmann, N., Shimbo, D., Burg, M., and Davidson, K. W. 2010. “Aspirin Adherence, Aspirin Dosage, and C-Reactive Protein in the First 3 Months after Acute Coronary Syndrome,” *American Journal of Cardiology* (106:8), pp. 1090–1094 (doi: 10.1016/j.amjcard.2010.06.018).
- Lazenby, M., Dixon, J., Bai, M., and McCorkle, R. 2014. “Comparing the Distress Thermometer (DT) with the Patient Health Questionnaire (PHQ)-2 for Screening for Possible Cases of Depression among Patients Newly Diagnosed with Advanced Cancer.,” *Palliative & Supportive Care* (12:1), pp. 63–8 (doi: 10.1017/S1478951513000394).
- Lee, J. K., Grace, K. A., and Taylor, A. J. 2006. “Effect of a Pharmacy Care Program on Medication Adherence and Persistence, Blood pressure, and Low-Density Lipoprotein Cholesterol: A Randomized Controlled Trial.,” *JAMA : The Journal of the American Medical Association* (296:21), pp. 2563–2571 (doi: 10.1016/S0145-4145(08)05023-5).
- Locklear, T., Miriovsky, B. J., Willig, J. H., Staman, K., Bhavsar, N., Weinfurt, K., and Abernethy, A. 2014. “Strategies for Overcoming Barriers to the Implementation of Patient--Reported Outcomes Measures.,”
- Lutfey, K. E., and Wishner, W. J. 1999. “Beyond ‘Compliance’ Is ‘Adherence’: Improving the Prospect of Diabetes Care,” *Diabetes Care*, pp. 635–639 (doi: 10.2337/diacare.22.4.635).
- Mayers, A. G., Grabau, E. A. S., Campbell, C., and Baldwin, D. S. 2009. “Subjective Sleep, Depression and Anxiety: Inter-Relationships in a Non-Clinical Sample,” *Human Psychopharmacology* (24:6), pp. 495–501 (doi: 10.1002/hup.1041).
- McKinley, S., Coote, K., and Stein-Parbury, J. 2003. “Development and Testing of a Faces Scale for the Assessment of Anxiety in Critically Ill Patients,” *Journal of Advanced Nursing* (41:1), pp. 73–79 (doi: 10.1046/j.1365-2648.2003.02508.x).
- McLennon, S. M., Bakas, T., Jessup, N. M., Habermann, B., and Weaver, M. T. 2014. “Task Difficulty and Life Changes among Stroke Family Caregivers: Relationship to Depressive Symptoms.,” *Archives of physical medicine and rehabilitation* (95:12), Elsevier Ltd, pp. 2484–90 (doi: 10.1016/j.apmr.2014.04.028).
- Min, Y. H., Lee, J. W., Shin, Y. W., Jo, M. W., Sohn, G., Lee, J. H., Lee, G., Jung, K. H., Sung, J., Ko, B. S., Yu, J. H., Kim, H. J., Son, B. H., and Ahn, S. H. 2014. “Daily Collection of Self-Reporting Sleep

- Disturbance Data via a Smartphone App in Breast Cancer Patients Receiving Chemotherapy: A Feasibility Study,” *Journal of Medical Internet Research* (16:5) (doi: 10.2196/jmir.3421).
- Mitchell, A. J., and Coyne, J. C. 2007. “Do Ultra-Short Screening Instruments Accurately Detect Depression in Primary Care? A Pooled Analysis and Meta-Analysis of 22 Studies.,” *The British journal of general practice : the journal of the Royal College of General Practitioners* (57:535), pp. 144–51 (available at <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2034175&tool=pmcentrez&rendertype=abstract>).
- Odinot, G., and Wolters, G. 2006. “Repeated Recall, Retention Interval and the Accuracy–Confidence Relation in Eyewitness Memory,” *Applied Cognitive Psychology* (20:7), pp. 973–985 (doi: 10.1002/acp.1263).
- Paes, A. H. P., Barker, A., and Soe-Agnie, C. J. 1997. “Impact of Dosage Frequency on Patient Compliance,” *Diabetes Care* (20:10), pp. 1512–1517 (doi: 10.2337/diacare.20.10.1512).
- Pasquini, M., and Biondi, M. 2007. “Depression in Cancer Patients: A Critical Review.,” *Clinical Practice and Epidemiology in Mental Health : CP & EMH* (3:2), p. 2 (doi: 10.1186/1745-0179-3-2).
- Peytchev, a., and Hill, C. a. 2009. “Experiments in Mobile Web Survey Design: Similarities to Other Modes and Unique Considerations,” *Social Science Computer Review* (28:3), pp. 319–335 (doi: 10.1177/0894439309353037).
- Pozuelo, L., Tesar, G., Zhang, J., Penn, M., Franco, K., and Jiang, W. 2009. “Depression and Heart Disease: What Do We Know, and Where Are We Headed?,” *Cleveland Clinic Journal of Medicine*, pp. 59–70 (doi: 10.3949/ccjm.75a.08011).
- Reid, S. C., Kauer, S. D., Dudgeon, P., Sanci, L. A., Shrier, L. A., and Patton, G. C. 2009. “A Mobile Phone Program to Track Young People’s Experiences of Mood, Stress and Coping,” *Social Psychiatry and Psychiatric Epidemiology* (44:6), pp. 501–507 (doi: 10.1007/s00127-008-0455-5).
- Robinson, J. H., Callister, L. C., Berry, J. a, and Dearing, K. a. 2008. “Patient-Centered Care and Adherence: Definitions and Applications to Improve Outcomes.,” *Journal of the American Academy of Nurse Practitioners* (20:12), pp. 600–7 (doi: 10.1111/j.1745-7599.2008.00360.x).
- Ruddy, K., Mayer, E., and Partridge, A. 2009. “Patient Adherence and Persistence with Oral Anticancer Treatment.,” *CA: A Cancer Journal for Clinicians* (59:1), pp. 56–66 (doi: 10.3322/caac.20004).
- Sandman, L., Granger, B. B., Ekman, I., and Munthe, C. 2012. “Adherence, Shared Decision-Making and Patient Autonomy,” *Medicine, Health Care and Philosophy* (15:2), pp. 115–127 (doi: 10.1007/s11019-011-9336-x).
- Scholle, S. H., Haskett, R. F., Hanusa, B. H., Pincus, H. A., and Kupfer, D. J. 2003. “Addressing Depression in Obstetrics/Gynecology Practice,” *General Hospital Psychiatry* (25:2), pp. 83–90 (doi: 10.1016/S0163-8343(03)00006-9).
- Seligman, M. E. P., Walker, E. F., and Rosenhan, D. . 2001. *Abnormal psychology, (4th ed)*, New York: W. W. Norton & Company.
- Snyder, C. F., Jensen, R., Courtin, S. O., Wu, A. W., Aaronson, N., Brundage, M., Gotay, C., Hodge, J., Hynes, D., Jones, J. B., Snyder, C., Wasson, J., Wu, A., Yount, S., and Zebrack, B. 2009. “PatientViewpoint: A Website for Patient-Reported Outcomes Assessment,” *Quality of Life Research* (18:7), pp. 793–800 (doi: 10.1007/s11136-009-9497-8).
- Sokol, M. C. M., McGuigan, K. K. a., Verbrugge, R. R. R., and Epstein, R. R. S. 2005. “Impact of Medication Adherence on Hospitalization Risk and Healthcare Cost,” *Med Care* (43:6), pp. 521–530 (doi: 10.1097/01.mlr.0000163641.86870.af).
- Stommel, M., Given, B. A., and Given, C. W. 2002. “Depression and Functional Status as Predictors of Death among Cancer Patients,” *Cancer* (94:10), pp. 2719–2727 (doi: 10.1002/cncr.10533).
- Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., and Hufford, M. R. 2002. “Patient Non-Compliance with Paper Diaries.,” *BMJ (Clinical Research ed.)* (324:7347), pp. 1193–1194 (doi: 10.1136/bmj.324.7347.1193).
- Strandbygaard, U., Thomsen, S. F., and Backer, V. 2010. “A Daily SMS Reminder Increases Adherence to Asthma Treatment: A Three-Month Follow-Up Study,” *Respiratory Medicine* (104:2), Elsevier Ltd, pp. 166–171 (doi: 10.1016/j.rmed.2009.10.003).
- Street, R. L., Makoul, G., Arora, N. K., and Epstein, R. M. 2009. “How Does Communication Heal? Pathways Linking Clinician-Patient Communication to Health Outcomes,” *Patient Education and Counseling* (74:3), pp. 295–301 (doi: 10.1016/j.pec.2008.11.015).

- Trask, P. C., Paterson, A. G., Hayasaka, S., Dunn, R. L., Riba, M., and Johnson, T. 2001. "Psychosocial Characteristics of Individuals with Non-Stage IV Melanoma," *Journal of Clinical Oncology* (19:11), pp. 2844–2850.
- Tu, C., Hsu, M.-C., Chi, S.-C., Lin, H.-Y., and Yen, Y.-C. 2014. "Routine Depression Screening and Diagnosing Strategy for Cancer Inpatients.," *Psycho-Oncology* (23:9), pp. 1057–67 (doi: 10.1002/pon.3547).
- Turnbull Macdonald, G. C., Baldassarre, F., Brown, P., Hatton-Bauer, J., Li, M., Green, E., and Lebel, S. 2012. "Psychosocial Care for Cancer: A Framework to Guide Practice, and Actionable Recommendations for Ontario," *Current Oncology* (19:4), pp. 209–216 (doi: 10.3747/co.19.981).
- Vahdaninia, M., Omidvari, S., and Montazeri, A. 2010. "What Do Predict Anxiety and Depression in Breast Cancer Patients? A Follow-Up Study," *Social Psychiatry and Psychiatric Epidemiology* (45:3), pp. 355–361 (doi: 10.1007/s00127-009-0068-7).
- Waterhouse, D. M., Calzone, K. A., Mele, C., and Brenner, D. E. 1993. "Adherence to Oral Tamoxifen: A Comparison of Patient Self-Report, Pill Counts, and Microelectronic Monitoring," *Journal of Clinical Oncology* (11:6), pp. 1189–1197.
- Williams, G. C., Freedman, Z. R., and Deci, E. L. 1998. "Supporting Autonomy to Motivate Patients with Diabetes for Glucose Control," *Diabetes Care* (21:10), pp. 1644–1651 (doi: 10.2337/diacare.21.10.1644).
- Williams, G. C., Gagné, M., Ryan, R. M., and Deci, E. L. 2002. "Facilitating Autonomous Motivation for Smoking Cessation.," *Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association* (21:1), pp. 40–50 (doi: 10.1037/0278-6133.21.1.40).
- Windschitl, P. D. 1996. "Memory for Faces: Evidence of Retrieval-Based Impairment.," *Journal of Experimental Psychology. Learning, Memory, and Cognition* (22:5), pp. 1101–1122 (doi: 10.1037/0278-7393.22.5.1101).
- World Health Organization. 2010. "HIV Drug Resistance Early Warning Indicators: World Health Organization Indicators to Monitor HIV Drug Resistance Prevention at Antiretroviral Treatment Sites.,"