

博士論文

**Capturing Real-Time Psycho-Behavioral Data in a
Natural Environment: the Utility of Ecological
Momentary Assessment**

(日常生活下における実時間での心理・行動データの収集：
エコロジカルモメンタリーアセスメントの有用性)

Jinhyuk Kim

金 鎮赫

**Department of Physical and Health Education
Graduate School of Education
The University of Tokyo**

2013

ACKNOWLEDGMENTS

I am deeply indebted to the many people whose help has made this thesis possible.

My special thanks go to Professor Yoshiharu Yamamoto for his constructive comments and enthusiastic encouragement. This thesis was inspired by his creative ideas and could not have been accomplished without his patience and guidance. His commitment to scientific endeavor is a model that I wish to emulate in my future research activities.

My heartfelt thanks also go to Dr. Hiroe Kikuchi. A number of informal meetings and discussions with her helped me identify interesting and useful approaches for this thesis. Without her basic and continuous instruction on how to perform research, I could not have completed this work.

Dr. Toru Nakamura gave me his firm support and encouragement in my doctor's course. I cannot thank him enough for his kind introduction to data analysis and a number of valuable discussions especially regarding the results of the studies introduced in parts 3 and 4 and how to write these results up.

Professor Tsukasa Sasaki and Professor Kazuhiro Yoshiuchi collaborated with me on the studies in parts 3 and 4. I would like to thank them for their generous help and many valuable comments, which developed and enhanced the quality of this thesis.

My thanks also go to the undergraduates and office workers at the University of Tokyo who kindly

participated in my experiments. I would also like to thank Dr. Naoko Aoyagi and Dr. Rika Nakahara for providing their data on adolescents and patients with major depressive disorder, respectively. Without the support and contributions of all these individuals, the collection of data for this thesis would have been impossible.

I am also grateful to Dr. Stephanie Coop for her assistance in improving the English expression in this thesis.

Professor Yoshiteru Mutoh, Professor Takashi Eto, Professor Gentaro Taga, Professor Daichi Nozaki, Professor Fumiharu Togo, Professor Kenji Morita, Professor Kentaro Yamanaka, Professor Masaya Hirashima, and Dr. Hyuntae Park provided instruction, support, and comments on my studies in many research meetings. I would like to express my sincere thanks to all of them.

My gratitude also goes to all members of the Department of Physical and Health Education, Graduate School of Education, The University of Tokyo.

Lastly, I would like to thank my family for supporting my life studying abroad.

Tokyo, Japan

October 2013

CONTENTS

| | |
|--|-----------|
| ACKNOWLEDGMENTS | i |
| PART 1: GENERAL INTRODUCTION | 1 |
| 1.1. Introduction to ecological momentary assessment (EMA)..... | 1 |
| 1.1.1. History and background of EMA | 3 |
| 1.1.2. Use of EMA in various studies..... | 6 |
| 1.2. Methodology of EMA..... | 8 |
| 1.2.1. EMA designs | 8 |
| 1.2.2. Devices used for collection of EMA data..... | 11 |
| 1.3. Methodological considerations in EMA | 14 |
| 1.4. Analysis of EMA data | 17 |
| 1.4.1. Multilevel modeling | 17 |
| 1.4.2. Model selection | 20 |
| 1.5. Research questions and aims of thesis | 21 |
| PART 2: MOMENTARY VERSUS RECONSTRUCTED FATIGUE AND MOOD..... | 24 |
| 2.1. Introduction..... | 24 |
| 2.2. Method | 27 |
| 2.2.1. Subjects | 27 |
| 2.2.2. Materials and procedure | 27 |
| 2.2.3. Measurements..... | 30 |
| 2.2.4. Data analysis..... | 32 |
| 2.2.5. Statistical analysis | 33 |
| 2.3. Results..... | 36 |
| 2.3.1. Subject characteristics | 36 |
| 2.3.2. Recording profiles | 36 |
| 2.3.3. Matching profiles of reconstructed recordings and momentary recordings | 36 |
| 2.3.4. Differences of time information between reconstructed recordings and momentary recordings | 37 |
| 2.3.5. Correlations and differences between reconstructed recordings and momentary recordings.. | 39 |
| 2.3.6. Effect of situational information on differences between reconstructed recordings and momentary recordings..... | 39 |
| 2.4. Discussion..... | 41 |
| PART 3: CO-VARIATION OF DEPRESSIVE MOOD AND LOCOMOTOR DYNAMICS | 45 |

| | |
|--|-----------|
| 3.1. Introduction..... | 45 |
| 3.2. Methods..... | 48 |
| 3.2.1. Subjects | 48 |
| 3.2.2. Materials, procedures, and measurements..... | 48 |
| 3.2.3. Data analysis..... | 51 |
| 3.2.4. Statistical analysis | 54 |
| 3.3. Results..... | 58 |
| 3.3.1. Recording profiles | 58 |
| 3.3.2. Convergent associations among symptoms..... | 59 |
| 3.3.3. Concurrent associations of depressive mood with local statistics of locomotor activity | 59 |
| 3.3.4. Cross validation for association of depressive mood with local statistics of locomotor activity | 63 |
| 3.4. Discussion..... | 65 |
| PART 4: CONTINUOUS ESTIMATION OF DEPRESSIVE MOOD FROM LOCOMOTOR DYNAMICS..... | 69 |
| 4.1. Introduction..... | 69 |
| 4.2. Methods..... | 71 |
| 4.2.1. Subjects | 71 |
| 4.2.2. Materials, procedures, and measurements..... | 72 |
| 4.2.3. Data analysis..... | 75 |
| 4.2.4. Statistical analysis | 75 |
| 4.3. Results..... | 78 |
| 4.3.1. Recording profiles | 78 |
| 4.3.2. Determination of statistical model for depressive mood | 79 |
| 4.3.3. Group differences in associations between depressive mood and locomotor activity | 81 |
| 4.3.4. Cross-validation of statistical model across groups | 82 |
| 4.4. Discussion..... | 83 |
| PART 5: GENERAL DISCUSSION..... | 88 |
| 5.1. Summary of findings..... | 88 |
| 5.2. Implication of findings..... | 91 |
| 5.2.1. Utility of EMA in field of psychiatry | 91 |
| 5.2.2. Using EMA with information and communication technologies | 92 |
| 5.3. Perspectives for future study..... | 94 |
| 5.4. Conclusion | 96 |

REFERENCES..... 97

SUPPLEMENTARY INFORMATION..... 106

PART 1: GENERAL INTRODUCTION

1.1. Introduction to ecological momentary assessment (EMA)

Self-reports are frequently used in the social and behavioral sciences to collect data on human behaviors and experiences [1]. Studies in these fields commonly ask subjects to report retrospectively about their own behaviors and emotions, or other outcomes. However, retrospective self-report data are often distorted by “recall bias” [2] and typically collected in laboratory environments by single measurements through questionnaires or interviews. Thus, our typical reliance on retrospective self-reports seriously limits our ability to accurately assess and evaluate subjects’ behavioral and emotional experiences without recall bias in real-world environments and misses the dynamics of experiences that change from moment to moment. EMA, an alternative to traditional self-reports, involves the repeated collection of self-reported information about a person’s momentary experiences in natural environments [2]. The key purposes of the EMA approach are to minimize recall bias, maximize ecological validity, and document variations over time [2,3].

Momentary assessments

EMA was developed as a means of acquiring self-reported data with less distortion from recall bias than that found in retrospective self-reports. Research on autobiographical memory [4] indicates that memory can be substantially inaccurate and can introduce systematic bias (cf. natural forgetting over time) with regard to experiences that are irregular. In other words, we often recall what has usually happened rather than what we have actually experienced. In addition, research using retrospective self-report inquiries asks subjects not only to recall but also to summarize their experiences. For example, when subjects report their experiences over the preceding week or month, recall will be

heavily affected by the most salient events during that period [2]. Those events that are more easily recalled, typically peak or most recent experiences, will bias the retrospective self-report. These properties inherent to retrospective self-reports, particularly summary reports over longer time periods, may thus give us a biased view of individuals' experiences. In contrast, people can provide accurate reports of their immediate experiences [1,5]. Therefore, EMA techniques circumvent recall bias by assessing phenomena through instantaneous reports of immediate experiences.

Real-world assessments

Most current researchers studying human behavior recognize that many experiences can be affected by the immediate environment, influencing the results of laboratory studies because the laboratory is not a typical environment. For example, a study on the phenomenon known as “white coat syndrome” has reported that blood pressure is elevated in a doctor's office but not in the real world [6]. Thus, for assessed experiences to be representative, the experiences have to be sampled in the contexts in which they naturally occur. EMA emphasizes ecological validity by collecting data in the real world, ensuring that the data represent our everyday life.

Frequent assessments

Frequent/repeated assessments of experiences through EMA can help investigate the dynamic nature of experiences themselves. This cannot be obtained by single measurements through retrospective self-reports. Researchers currently seek data with richer temporal resolution to test dynamic correlations among variables (e.g., psychological and physiological states) over time. For example, EMA studies use temporal resolution to focus on the within-individual changes in psychological experiences and physiological states over time, in order to understand how certain psychological

variables precede/follow other physiological variables. Momentary fluctuations in the experiences and the interplay among their various components monitored by EMA are useful for better understanding the interactions among such variables.

1.1.1. History and background of EMA

EMA is not a single method but rather a collection of methods that share the common characteristics of EMA described above. Although the design of these methods, such as the technologies or skills used, the targets of assessment, and the schedules for data collection, vary, the EMA approach aims to bring these diverse methodologies together under a common framework. The characteristics of EMA have been developed on the basis of several historical traditions, including diaries, time budget studies, behavioral observation, self-monitoring, the experience sampling method (ESM), and ambulatory assessment (AA) [3,7].

Diaries and time budget studies

Diaries have long been used to collect research data about subjects' daily experiences [7]. There are various types of diaries, from those involving free description to structured types, and this method has been used in a wide range of research fields, including behavioral research, e.g., collecting self-reports on coping behaviors. At times, diaries are used to record the results of monitoring of physiological states such as blood pressure and salivary cortisol in clinical research.

Studies using the time budget method, which is a structured type of diary, have been concerned with describing how people experience the activities in their lives and how they spend their time [8]. This method has traditionally been used in quantitative research, but could be used in qualitative research to measure past behavior, attitudes, and emotions [9,10,11]. For example, the day

reconstruction method (DRM) involves subjects first recalling the previous day by constructing a diary consisting of a sequence of behavioral episodes with time information, and then recording their experience of each behavioral episode and any associated symptoms [9,11]. In addition, event history calendars (EHCs) have also been used to address the reliability and validity of retrospective data collection and collect retrospective data about events and life transitions over short and long time periods [10]. Time budget studies including DRM and EHCs attempt to solve concerns about the reliability and validity of the retrospective data collected. These techniques are thought to be well-structured approaches to recording that facilitate subjects' recall of past events by using their own past experiences as cues for reconstruction.

Behavioral observation and self-monitoring

Direct description and coding of subjects' behavior has been used by sociologists and psychologists interested in the effect of particular contexts on behavior [12]. Behavioral observation allows us to understand how subjects allocate their time and what activities they engage in throughout the day in the natural environment. In addition, these methods permit a more objective evaluation of particular behavior than self-reports. However, these methods tend to be burdensome and limit the range of behaviors and the duration of observation in some contexts [7]. For example, they cannot be used to observe behaviors such as sexual behavior and marital conflict, and if used for extended periods they would constitute an invasion of privacy.

Self-monitoring (i.e., behavioral observation alone) of particular behaviors or experiences has a long history in behavioral research. This method involves two component responses [13]. One is that subjects discriminate or notice the occurrence of the target event (e.g., action, thought, or feeling) and the other is that subjects produce additional information about the antecedents and contexts of the

target event (e.g., having a meal [14]). Self-monitoring often aims to capture behaviors targeted for change in behavioral treatment [15,16] and this approach can sometimes help to change the target behavior [16].

Experience sampling method

ESM is a method for collecting information about the context and content of the daily life of individuals [17]. The advantage of ESM is its ability to capture daily life as it is directly perceived from one moment to the next, making it possible to examine fluctuations in the external context and in the content of individuals' subjective experiences. The method achieves this by asking individuals to immediately record responses at several random points throughout the day [17,18]. Self-reports with ESM mainly concentrate on ecological validity, and the problem of recall bias is automatically solved at the same time. In the period before the introduction of devices such as palmtop computers, ESM researchers asked subjects to carry pagers so that they could be prompted to record their experiences on paper for later collection and analysis [7].

Ambulatory assessment

Enabled by technological developments, AA has become an essential research tool over the past two decades. AA covers a wide range of assessment methods that are used to study subjects in their natural environment via continuous or near-continuous recording [19]. It is most frequently used to obtain physiological or biological data. Ambulatory monitoring of cardiovascular function, using portable cardiac monitors, has been used for several decades as a tool for understanding the association between experiences and cardiovascular health [20]. Recent developments have expanded physiological monitoring to other parameters, such as physical activity [21,22,23,24,25,26,27],

hypothalamic-pituitary-adrenal axis activity [28,29,30,31], blood glucose [32], skin temperature [33], pulmonary function [34], and others. In addition, technological developments have enabled automated EMA assessment of behaviors (e.g., pill taking) [35] and physical environment (e.g., air sampling) [36]. Studies using AA have the advantage, relative to laboratory research, of ecological validity which reflects subjects' environmental and interpersonal factors that change over time.

1.1.2. Use of EMA in various studies

Since the term EMA was proposed in 1994, this approach has been actively used to collect data on a very wide range of behaviors, experiences, and states. A review of diary studies [37] showed that there are large groups of studies using EMA that deal with pain, mood and affect, anxiety, eating, sleep, alcohol consumption, and psycho-physiological states. EMA studies address all these types of data and also include studies of depression [38,39,40], psychological stress [41,42], self-esteem [43], diet [44], self-reported physical activity [45,46], smoking [47,48,49], sexual behavior [50], compulsive buying [51], social interaction [38,52], work activity and satisfaction [42,53], diabetes management [52,54], effects of medication [55,56], asthma [34,57], allergies [36,58], tinnitus [59], and so on. In addition, clinical disorders studied with EMA include the full range of psychopathology, such as addictive disorders [60,61], gastrointestinal disorders [62], sexual dysfunction [63], eating disorders [64,65,66], attention deficit hyperactivity disorder (ADHD) [67,68], mood dysregulation [69], anxiety disorders [70,71,72], depressive disorders [73,74,75,76], bipolar disorder [77], and schizophrenia [78,79,80]. Furthermore, EMA is widely used to evaluate treatment and intervention of crucial health-related behaviors in health psychology and behavioral medicine, such as coping with illness and treatment [64,81], medication compliance [82,83], and exercise [84,85]. EMA techniques are now used in most fields of social and behavioral sciences and are continuing to spread to other areas of

sciences.

1.2. Methodology of EMA

When researchers use retrospective assessments such as trait questionnaires, they assume that the assessment captures what subjects believe or remember in a single assessment [86]. Thus, it may be unimportant when the assessment is made. However, EMA is not a series of fixed protocols, but an approach to capturing moments or periods of real-time data by sampling it repeatedly over time, raising the issue of how to ensure that the moments or periods assessed are representative of the data [7]. A well-designed EMA sampling scheme can precisely provide almost complete coverage of behaviors and experiences in an individual's life over time as well as yield new insights into them.

1.2.1. EMA designs

EMA sampling can be roughly divided into time-based and event-based assessment [2]. Time-based assessment typically aims to characterize continuous experience, e.g., observing how mood varies over time, without a predefined focus on particular discrete events. Event-based assessment does not aim to characterize subjects' entire experience, but rather to monitor particular discrete events or episodes in their lives, e.g., smoking episodes [87], and focus on data collection around these events. Also, while some EMA designs use a single EMA schedule or strategy, it is often useful or essential to combine them [3,7].

Time-based assessment

Time-based assessment is particularly well suited for clinical phenomena that vary continuously (e.g., mood states, pain, social interaction) and are not easily conceptualized in an episode. While some physiological states (e.g., actigraphy, heart rate) can be monitored continuously, in the more typical case, EMA protocols rely on time-based assessment. There are many types of time-based assessment

strategies which vary in schedule, frequency, and timing.

A variety of time-based assessment schedules are used in EMA. Some studies have adopted a fixed-interval scheduling of assessment. The assessment interval has varied substantially, ranging from once or several times a day to intensive schedules. For example, subjects complete an assessment once daily, usually near the end of the day before going to bed. Assessments completed several times a day are conducted in the morning, afternoon, and evening, or at certain times of the day (e.g., every four hours). Also, it is common that ambulatory monitoring of blood pressure is automatically completed every 30 to 45 min [88]. However, this fixed-assessment provides predictability for subjects and they may come to anticipate the assessments in ways that change their experiences. On the other hand, when the time blocks are ambiguously defined (e.g., morning), subjects implicitly have great discretion in the timing of assessments. For example, subjects might be reminded to complete their pain assessments when they are in pain, thus biasing the sample of pain.

To avoid this potential bias in fixed-interval time-based assessment, EMA studies generally use a random sampling strategy, characterizing subject experience through a series of randomly scheduled assessments to achieve representation of the subject's experience. Many EMA researchers have used stratified random schedules in which the day is divided into blocks of time and an assessment is placed at random within each block [89], ensuring that assessments are evenly performed throughout the day.

Although the frequency of assessments depends on the purpose of the study and the target behaviors or experiences studied, it is necessary to consider subject respondent load. Thus, EMA studies usually assess subjects 3 to 5 times per day [3]. In particular cases, some studies have made as many as 20 or more assessments a day (e.g., daily fluctuation in physiological states) [88,90,91,92]. In addition, assessment frequency may be varied to collect more assessments in certain time blocks if

those are of particular interest (e.g., alcohol consumption at night time) [93]. Furthermore, some studies have limited assessment to a certain range of time (e.g., waking time) according to the study goals [78].

Event-based assessment

Event-based assessment is suitable when the research interest is conceptualized by discrete events or episodes. This approach asks subjects to record the occurrence of a predefined event. For example, subjects might be instructed to complete an assessment when they feel a tension-type headache [94], or smoke a cigarette [47,48,49]. Because, in most cases, the subjects themselves initiate an assessment when the predefined event has occurred, the event must be clearly defined. If subjects were required, for example, to make a record every time they engaged in a “social interaction,” it would be confusing for them because the definition is not concrete enough. In this case, the event could be redefined as a “social interaction lasting more than 10 min” [95]. In addition, the researcher may wish to collect data not only on the frequency of a predefined event but also on other experiences that occurred simultaneously (e.g., mood states, location). If the event frequency is too high, however, it may not be possible for subjects to assess each event. In these cases, recording of other experiences could be restricted to larger event units (e.g., a bout of drinking), or randomly sampled by time-based EMA [96].

Combination designs

When researchers focus on momentary states and assess them intermittently, the assessments do not provide complete coverage of the subjects’ whole experiences. A combination of time- and event-based assessments increases the chance of recording the subjects’ experiences more accurately.

In addition, although data obtained once a day may miss important sources of dynamic variation in behavior, daily diaries can be used as a supplementary method for capturing data on highly memorable and slow-moving targets. A combination of daily diaries and time-based (or event-based) assessment may also be useful when researchers especially seek to record subjects' retrospective cognitive processes along with their momentary responses throughout the day.

Combined sampling approaches are a fruitful way of testing particular hypotheses. Event-based assessment is often used to collect data about the circumstances or antecedents that are associated with a target event. For example, to understand the association between positive/negative affect and smoking, we first collect data on smoking occasions using event-based assessment. However, it is difficult to use event-based assessment to sample nonsmoking occasions as a control for the smoking occasions since nonsmoking episodes cannot be tracked as an event [96]. By incorporating time-based assessment for capturing nonsmoking occasions, we can also document the antecedent or following smoking occasions.

Combining different schedules of time-based assessments can also be useful. For example, alcohol consumption as well as assessing throughout the day, an assessment could also be scheduled each morning to ask about hangovers from the previous night's drinking [97]. This reminds us that careful consideration about the arrangement and scheduling of assessments need to be driven by the research questions and by the natural time course of the target behavior.

1.2.2. Devices used for collection of EMA data

EMA studies select a device for recording and collecting data according to study purpose and situation. There are various types of devices, ranging from basic and simple methods (e.g., paper-and-pencil diaries) to state-of-the-art technologies (e.g., smartphones). The devices are generally used to provide

Table 1-1. Devices used for collection of EMA data

| Device/technology | Functions | | | | | |
|---------------------------------------|-----------------|----------------|-------------------------|-------------------------|-------------------------|--------------------|
| | Purpose of use | Self-report | Time-stamp ¹ | Manage scheduling | | |
| | | | | predefined ² | responsive ³ | portable |
| Paper-and-pencil diary | R ⁴ | Y ⁵ | × | × | × | ○ |
| Tape recorder | R | Y | × | × | × | ○ |
| Web | R | Y | ○ | × | × | ○ ⁶ /× |
| Physiological monitor | R | N ⁷ | ○ | ○/× | × | ○ |
| Event recorder ⁹ | R | N | ○ | × | × | ○ |
| Biological sample ¹⁰ | R | N | × | × | × | ○ |
| Beeper | P ¹¹ | N | × | ○ | × | ○ |
| Watch | P | N | × | ○ | × | ○ |
| Watch-type computer | P + R | Y | ○ | ○ | × | ○ |
| IVR ¹² with incoming calls | R | Y | ○ | ○ | × | ○ ¹³ /× |
| IVR with outgoing calls | P + R | Y | ○ | ○ | ○ | ○ ⁹ /× |
| Laptop computer | P + R | Y | ○ | ○ | ○ | × |
| Tablet computer | P + R | Y | ○ | ○ | ○ | × |
| Palmtop computer | P + R | Y | ○ | ○ | ○ | ○ |
| Smartphone | P + R | Y | ○ | ○ | ○ | ○ |

○ Function is provided.

× Function is not provided.

1 Data records are stamped with time information.

2 For example, random scheduling that can be predefined.

3 Allows for change in scheduling in response to certain data received.

4 Recording (R): Device/technology can record data.

5 Yes (Y): Device/technology can record self-reported data.

6 If smartphone used.

7 No (N): Device/technology cannot record self-reported data.

8 Usually handles only fixed time schedules.

9 For example, instrumented medication containers.

10 Devices for collecting biological samples (e.g., blood pressure, saliva).

11 Prompting (P): Device/technology can prompt subjects to record data.

12 Interactive voice response (IVR): Automated telephone systems that use normal phone calls to prompt subjects with prerecorded questions and receive responses via telephone keypad input.

13 If cell phone used.

the following functions. First, the devices are used to present questionnaires and available response options and to save these collected data with time information (i.e., time-stamped data). In addition, when using time-based assessment, the devices prompt the subjects to record with signals (e.g., by “beeping”) at the appropriate times [7]. The type of devices used in EMA studies and their detailed functions are summarized in Table 1-1 (modified from [7]).

1.3. Methodological considerations in EMA

There are potential methodological issues associated with the repeated collection of real-time data in EMA studies, because the data is recorded by subjects themselves in natural environments without the supervision of a researcher. Researchers have successfully applied a variety of assessment platforms to the collection of EMA data. As the implementation of EMA becomes more ubiquitous across psychological, social, and physiological variables, researchers will increasingly be able to take their research out of the lab and into the natural environment of their subjects' daily lives.

Compliance

In time-based assessment, the success of an EMA study depends on a high rate of compliance with the sampling scheme protocol. If subjects did not respond to a high proportion of prompts, researchers might have their doubts about the representativeness of the sampling and generalizability of the findings. Missing assessments have the potential to bias the obtained recordings of behavior and experience, if the missing data are nonrandom. For example, in a study on the intensity of fatigue, it is easy to imagine that subjects might ignore a prompt when they are exhausted. If so, the recordings of the level of fatigue for the period studied would be falsely underestimated. Thus, missing data are likely to be systematic and may be a threat to the validity of a study.

A related issue involves the possibility that subjects are not completing momentary recordings according to the protocol but are completing the recordings at a later time. An extreme example of this problem occurs when an entire day's reports are hoarded and backfilled. Also, even in this case, subjects may report that they have completed the diaries according to the protocol schedule. There is no way for researchers to know for sure when the recordings were actually made. A study [98] has confirmed these concerns about compliance with paper-and-pencil diaries. Using a specially

instrumented paper diary binder that enabled the researchers covertly monitor when diary was opened and closed, it was possible to document when diary entries were actually made and compare this with the time and date subjects entered in the written diaries. Across a 3-week reporting period, reported compliance was 90%. However, only 11% of entries were actually compliant based on the diary binder being open and having a completed diary entry for the corresponding time. Moreover, there was evidence of frequent hoarding of diaries across days. This result suggests that poor compliance is a major issue in research using paper-and-pencil diaries.

A solution to accurate recording of compliance is to use an electronic diary (ED) which enables data records to be stamped with time information in a manner that is not accessible to subjects. This function is available in a number of different EMA devices as described above (see Table 1-1). Although many EMA studies using ED report compliance rates ranging from 75%–90% [99], the ED is not an unconditional solution for achieving a high level of compliance, given that in some studies compliance rates have been as low as 50% [100]. It is necessary to train subjects on how to use the ED, and use a combination of well-structured protocols and questionnaires, feedback, incentives to participate, and other procedures [7].

For event-based assessment, there is an important limitation in which subjects' compliance with event recording is crucial to the validity of the data but often difficult to confirm. A study using carbon monoxide (CO) as a biochemical marker roughly confirmed that subjects are adequately compliant with ED records of cigarettes [101]. In addition, medication compliance is validated using objective measures of instrumented pill bottles [35]. Although only few studies on this issue exist so far, technological developments may improve compliance in some types of events. Researchers should carefully consider how a low level of compliance may affect the data.

Reactivity

Reactivity is defined as the possibility of behavior or experiences being affected by the research method itself. The EMA approach could be regarded as particularly vulnerable to reactivity because it involves repeated assessments. The course of subjects' daily lives may be interrupted by recording throughout the day, and this might cause distress that could bias the data. However, several studies have found little or no evidence of reactivity to EMA [100,102,103]. A well-designed study [104] also found no difference in trends over time with regard to pain and no systematic effect on intensity of pain, despite an increase in the number of scheduled assessments (i.e., 3, 6, or 12 times daily).

However, other studies have shown that reactivity emerges when subjects are trying to change the target behavior (e.g., undesirable target such as substance use) and when researchers ask subjects to record antecedents of specific events (e.g., recording a meal before eating it) [13]. If subjects are motivated to change their behavior and have a chance to reconsider the target event before recording it, reactivity could be enhanced. This suggests that EMA researchers need to be aware of reactivity under such conditions, although there is little evidence to date indicating strong reactivity effects.

1.4. Analysis of EMA data

EMA data usually have a hierarchical structure consisting of a large number of observations for each subject, with the number and timing of observations often varying between subjects. In addition, EMA data almost always have missing data, which cannot be analyzed using traditional techniques such as repeated measures analysis of variance (RM-ANOVA), because subjects cannot easily complete whole EMA recordings according to the protocol schedule. However, such data can be handled by the multilevel modeling approach [105,106], which is an extension of traditional regression models and has been recommended for the analysis of data with a hierarchical structure, including EMA data [107].

1.4.1. Multilevel modeling

Multilevel modeling can handle unbalanced data in which the number of the EMA recordings vary between individuals. In multilevel modeling, these within- and between-individual effects can be handled together in the same model by incorporating random effects into model coefficients, i.e., allowing the coefficients to vary across individuals. Although the multilevel model can be expressed as a single equation, it is easier to understand if it is initially presented as a set of equations that are separated within- and between-individual levels. In EMA analysis, the within-individual level (level 1) units are observations and the between-individual level (level 2) units are subjects.

Simplest multilevel model

The simplest multilevel model is the so-called null model. In this model, there is a dependent variable with no predictor, including within- and between-individual effects.

Level 1 equation (within-individual level):

$$Y_{ij} = \pi_{0i} + \varepsilon_{ij} \quad (1)$$

Level 2 equation (between-individual level):

$$\pi_{0i} = \gamma_{00} + \zeta_{0i} \quad (2)$$

Combined model:

$$Y_{ij} = \gamma_{00} + \zeta_{0i} + \varepsilon_{ij} \quad (3)$$

where Y_{ij} indicates the dependent variable at the j th observation for the i th subject; π_{0i} is the subject i 's intercept; γ_{00} is the average intercept across all subjects; and the random terms ζ_{0i} and ε_{ij} are the between- and within-individual residuals. The random term ζ_{0i} in (2) is included to account for that portion of the individual differences in the intercept π_{0i} , indicating that the intercept can vary across individuals.

Multilevel model with predictors

The null model above can be expanded by adding predictors into the level 1 equation (1). Hence, the specification of this model is as follows:

Level 1 equation (within-individual level):

$$Y_{ij} = \pi_{0i} + \sum_{k=1}^n \pi_{ki} X_{ij} + \varepsilon_{ij} \quad (4)$$

Level 2 equations (between-individual level):

$$\pi_{0i} = \gamma_{00} + \zeta_{0i} \quad (5)$$

$$\pi_{ki} = \gamma_{k0} + \zeta_{ki} \quad (k = 1, \dots, n) \quad (6)$$

Combined model:

$$Y_{ij} = \gamma_{00} + \sum_{k=1}^n \gamma_{k0} X_{ij} + \zeta_{0i} + \sum_{k=1}^n \zeta_{ki} X_{ij} + \varepsilon_{ij} \quad (7)$$

where X_{ij} is the predictor corresponding to the j th observations for the i th subject; π_{0i} and π_{ki} are the subject i 's intercept and coefficient (i.e., slope) of the predictor, respectively; γ_{00} is the average

intercept across all subjects; γ_{k0} is the average slope across all subjects; the random terms ζ_{0i} and ζ_{ki} are the between-individual residuals; and ε_{ij} is the within-individual residual. Multilevel models can consist of a linear combination of a subset of predictors X_{ij} in the above equations with the total number of n and their interactions. In addition, whether each predictor can vary across individuals can be determined by adding the random terms ζ_{0i} and ζ_{ki} in (5) and (6), respectively.

Multilevel model with a categorical variable

To examine the differences across between-individual-level variables (e.g., gender, occupation) in the relationship between the dependent variable and predictors, a categorical variable is added into the intercept and slopes.

Level 1 equation (within-individual level):

$$Y_{ij} = \pi_{0i} + \sum_{k=1}^n \pi_{ki} X_{ij} + \varepsilon_{ij} \quad (8)$$

Level 2 equations (between-individual level):

$$\pi_{0i} = \gamma_{00} + \gamma_{01} Z_i + \zeta_{0i} \quad (9)$$

$$\pi_{ki} = \gamma_{k0} + \gamma_{k1} Z_i + \zeta_{ki} \quad (k = 1, \dots, n) \quad (10)$$

Combined model:

$$Y_{ij} = \gamma_{00} + \gamma_{01} Z_i + \sum_{k=1}^n \gamma_{k0} X_{ij} + \sum_{k=1}^n \gamma_{k1} Z_i X_{ij} + \zeta_{0i} + \sum_{k=1}^n \zeta_{ki} X_{ij} + \varepsilon_{ij} \quad (11)$$

where Z_i is a dummy variable representing items of the categorical variable (e.g., male or female for gender) which the i th subject is assigned to, taking the value 0 for the referent and 1 for the other(s).

γ_{01} and γ_{k1} are determined by the categorical variable. Note that none of the between-individual-level variables are included in the level 1 equation because they are constant (have no variance) across the within-individual-level observations of the i th individual. If a categorical variable (e.g., subjects' activity, location and companion at each EMA observation) varies at the within-individual level, the

categorical variable is added to the right-hand side of (8).

1.4.2. Model selection

When selecting models in general statistics, if the fit of the simpler model to the data is worse than the fit of the more complex model incorporating predictors, the simpler model is rejected as inadequately representing the variation in the data. The goodness of fit of the models is compared by using two types of tests depending on the structure of the multilevel models. When comparing models with nested or inclusive relations, where all predictors used in one model were a subset of the predictors used in another model, the deviance statistic is tested [105]. This approach tests the difference in the $-2 \log$ -likelihood on the basis of the chi-square test with degrees of freedom equal to the difference in the number of model parameters to be estimated. If there is no significant difference in the $-2 \log$ -likelihood of two models, the simpler model will typically be preferred on the basis of the *principle of parsimony*. For the non-nested models, we use a statistical test based on the Akaike information criteria (AIC).

In addition, when we evaluate the goodness of fit for the multilevel models, we should consider the inclusion of a random effect into the model parameters [i.e., intercept and slope(s)] [105]. Whether the model is significantly improved by adding a random effect on the parameters is also tested with the deviance statistic.

1.5. Research questions and aims of thesis

Evaluation of the subjective symptoms of disease is the basis of its diagnoses and treatment, especially in cases of psychiatric and psychosomatic diseases. However, much of our evaluation of individuals' symptoms, including mood states, fatigue, and other emotional experiences, is based on single assessments using retrospective self-reports. Researchers and clinicians have indicated that retrospective self-reports are often inaccurate because recall of the past with retrospective self-reports is affected by processes more complex than natural forgetting [2,4]. Also, researchers and clinicians currently seek data with temporal variations to identify dynamic correlations among variables (e.g., psychological and physiological states) over time depending on the research or clinical purposes. Thus, new recording strategies such as EMA focus on self-reports of momentary experience, aiming to collect temporal variations in the momentary experience without recall bias.

As EMA is used in a wide range of research fields, including clinical settings, for capturing subjective symptoms, many studies to examine the utility of EMA have been conducted. These can be divided into two main categories: 1) studies that have attempted to demonstrate that EMA can minimize recall bias by comparing momentary symptoms with reconstructed symptoms, and 2) studies that have examined which factors are reflected in EMA recordings by investigating the associations between momentary symptoms and external objective criteria (e.g., physiological states). However, these studies have suffered from certain problems, which this thesis addresses, as explained below.

With regard to the first category, after EMA was first proposed, many studies demonstrated the strength of EMA in terms of the minimization of recall biases by comparing momentary (i.e., EMA) and retrospective self-reports [108,109,110]. However, systematic research has not rigorously conducted how the temporal variations in momentary symptoms compare to those in reconstructed

ones because the comparison of momentary and retrospective symptoms was made with their averaged observations across certain monitoring periods. Therefore, in part 2 of this thesis, I examine the differences and correlations between momentary symptoms (i.e., fatigue and mood states) recorded using EMA and those reconstructed using simultaneous DRM in healthy adults on a within-individual basis.

With regard to the second category, with the recent development of devices and techniques for real-time monitoring of externally objective processes, many studies have started to examine the utility of EMA for capturing subjective symptoms by comparing the within-individual temporal variations in momentary symptoms compared with simultaneously measured external criteria [28,29,30,31,34,88,91,111]. Although these studies focus on which external criteria are associated with momentary symptoms, the external criteria (e.g., blood pressure and salivary cortisol) were usually collected in a discontinuous and obstructive fashion and the possibility cannot be entirely excluded that the collection of the criteria themselves affected momentary symptoms. Therefore, in part 3 of this thesis, I first examine the variations in momentary symptoms by validating the associations among the self-reported symptoms measured simultaneously and then investigate psycho-behavioral correlations between the symptoms (especially, depressive mood) and spontaneous physical activity (i.e., locomotor activity) as the external objective criteria obtained continuously and un-obstructively in healthy humans.

In part 4, I focus on whether the same psycho-behavioral correlations are found also in patients with psychiatric disorders because showing such correlations in the patients is considered crucial in validating the real-time assessment for psychiatric disorders. In this part, therefore, I extend the study in part 3 and investigate psychobehavioral correlates, particularly the statistical associations between momentary depressive mood and behavioral dynamics measured objectively, in patients with major

depressive disorder (MDD).

The overall purpose of this thesis is to examine the utility of EMA as a method for capturing momentary symptoms in natural settings. This can give us basic insights into the understanding of self-reported symptoms and more effective strategies for collecting valid information to address important scientific questions related to them. It is my hope that this thesis will be a first step toward encouraging such insights and accelerating progress in diagnosis, treatment, prevention, and early detection of psychiatric disorders.

The studies in this thesis are based on the published and submitted papers listed below:

1. Kim J, Kikuchi H, Yamamoto Y (2013) Systematic comparison between ecological momentary assessment and day reconstruction method for fatigue and mood states in healthy adults. *Br J Health Psychol* 18: 155-167. (Basis of part 2)
2. Kim J, Nakamura T, Kikuchi H, Sasaki T, Yamamoto Y (2013) Co-variation of depressive mood and locomotor dynamics evaluated by ecological momentary assessment in healthy humans. *PLoS ONE* 8: e74979. (Basis of part 3)
3. Kim J, Nakamura T, Kikuchi H, Yoshiuchi K, Sasaki T, Yamamoto Y (in revision) Continuous estimation of depressive mood from spontaneous physical activity in major depressive disorder. (Basis of part 4)

PART 2: MOMENTARY VERSUS RECONSTRUCTED FATIGUE AND MOOD

2.1. Introduction

Self-report is the most common and potentially the best way to measure a person's subjective symptoms [1]. At the same time, it is also true that self-reports of behaviors and affect are strongly influenced by features of the study instrument, including study protocol, questionnaire format, and context. Many studies by psychologists and survey methodologists have been devoted to understanding the nature of self-reports and to improving the quality of data collection without methodological biases [112]. In this study, I discuss how two different assessment techniques – ecological momentary assessment (EMA) and the day reconstruction method (DRM), both of which were mainly designed to reduce the influence of recall bias [113,114,115,116] – record physical symptoms and mood states in healthy adults.

EMA is a method to acquire self-reported information at the moment. Its recording of instantaneous states of feeling allows the researcher to bypass problems of recall bias [2,3]. A paper-and-pencil diary was initially used for EMA, but it has been reported that the diary has the disadvantage of delayed data entry or even forward entry of data (i.e., “faked compliance”) [98]. Computerized EMA using computers as an electronic diary (ED) has been developed to avoid faked compliance by registering input time to the device automatically. Symptom diaries derived from computerized EMA are generally regarded as the “gold standard”, but EMA data based on sparse sampling may not provide contiguous data in spite of high levels of respondent load [9,11,117]. To overcome this issue, DRM was more recently proposed as an alternative method. In DRM, subjects first recall and record a sequence of behavioral episodes including situational information (e.g.,

activity, location, and companion) they have experienced throughout the day, and then record their emotional experiences (e.g., happiness, tiredness, etc.) in all the behavioral episodes they have first recorded [9,11,118]. DRM can be measured at some later time (e.g., on the next day) and therefore not only imposes less respondent load but also does not disrupt daily activities. DRM may also be a more useful method when EMA is difficult to conduct due to, for example, dangerous or busy circumstances of work.

While both the EMA and DRM aim at overcoming the recall bias in self-reports of symptoms, so far a full systematic comparison of these methods has not been conducted. In earlier studies simultaneously performing EMA and DRM, the comparison of the results was not made with the same subjects or for the same study period [9,11]. In a recent study on the correspondence between EMA and DRM of the same subjects and for the same study period [119], EMA was performed using a paper-and-pencil diary and the time information of EMA recordings was not obtained objectively. In addition, changes over time of the observations by EMA and DRM could not be rigorously examined because the comparison was made with the averaged observations across certain monitoring periods (e.g., every few hours). Therefore, the purpose of this study was to investigate the correlations and differences on a within-individual basis between momentary assessments recorded with computerized EMA and reconstructed assessments recorded with DRM in the same subjects and for the same study period.

By systematically comparing EMA and DRM recordings at each observation, I was able to investigate the correspondence of time information and how well changes over time of reconstructed fatigue and mood states corresponded to momentary fatigue and mood states measured not only by the standard time-based EMA but also by a modified assessment, behavioral “episode-based EMA”. I also investigated whether reconstructed symptoms were systematically biased by certain situational

information which was originally intended for use as a help in reconstructing symptoms in DRM.

2.2. Method

2.2.1. Subjects

The subjects were recruited from the Department of Physical and Health Education at the University of Tokyo. All the subjects who participated in this study were given a full explanation of the purposes and potential risks of the study by well-trained researchers. Thirty-five healthy undergraduates applied to participate, and were all enrolled. All the subjects signed an institutionally approved informed consent form to participation.

2.2.2. Materials and procedure

Momentary fatigue and mood states: ecological momentary assessment

To record momentary fatigue and mood states, a watch-type computer (Ruputer, ECOLOG, 42 g; Seiko Instruments Inc., Tokyo, Japan) [94] was used as an ED. The computer was equipped with a screen measuring 20 × 30 mm and a joystick and buttons as input devices (Figure 2-1A). The subjects were fully instructed how to use the device and given manuals before the beginning of the study period. They also practiced manipulating the device until they became accustomed to its use.

The subjects wore the watch-type computer and recorded EMA questionnaires for three days scheduled by two assessment conditions: time-based EMA and episode-based EMA. For the time-based EMA, the subjects were prompted to record the EMA questionnaires for two days, every two hours with a beep as a signal. The signals were programmed to occur randomly within intervals of plus or minus 12 min, in order to minimize anticipatory effects. If the subjects did not enter a recording when the computer beeped, they were allowed twice to postpone input for 10 min. Recordings not made within 20 min were cancelled. The subjects were also asked to record the EMA questionnaires when they woke up and went to bed by choosing “waking up” or “going to bed” from

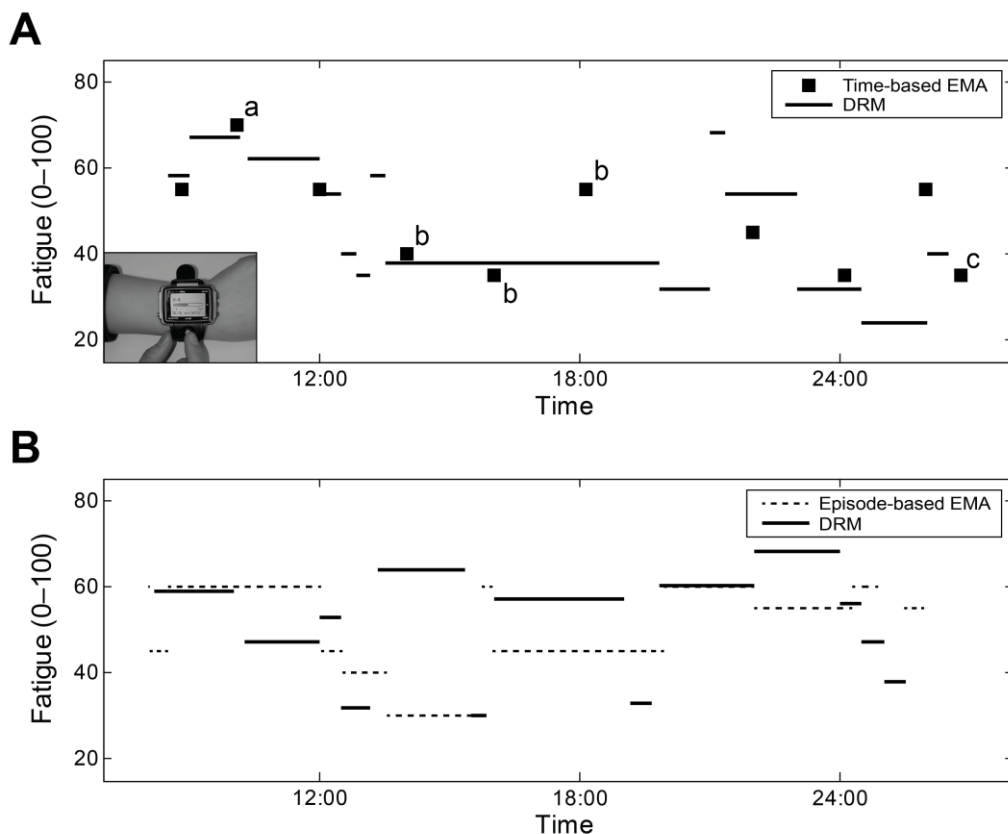


Figure 2-1. Example of a subject's fatigue intensity assessed by day reconstruction method (DRM) and ecological momentary assessment (EMA) for a day. Square in (A) indicates fatigue intensity recorded with time-based EMA at that moment approximately every two hours, and dashed line in (B) shows fatigue intensity recorded with episode-based EMA during one episode which had just finished (the time at the right end of the line) when switching from one episode to another. Solid line in (A) and (B) is for fatigue intensity recorded with DRM afterward according to series of behavioral episodes about the same days as when EMA was performed. The inset in the lower left of (A) represents the watch-type computer equipped with EMA software. Situational information was also recorded at each assessment (e.g., a: “studying”, b: “part-time job”, and c: “going to bed”).

the menu. After selecting the “going to bed” setting, the computers suspended the time-based recordings until the “waking up” setting was selected, to avoid disturbance of sleep.

Time-based EMA is generally used for EMA recording, however it may not assess the same point in time as DRM assesses. While subjects in the DRM assessment are asked to rate their situational information and any symptoms within the time span of each behavioral episode, time-based EMA evaluates them only at the prompted moment. Therefore, in the present study I also used a modified

EMA technique, episode-based EMA, in order to compare EMA with DRM within a comparable temporal framework. The subjects recorded fatigue and mood states with episode-based EMA in each behavioral episode (a time span with a specific behavior, e.g., studying with friends at school) for one day. The episode-based assessments were carried out by the subjects without a beep when one episode had occurred and just before the following episode. For example, the episode-based assessment was conducted when a subject had finished breakfast with his/her family and just before he/she left home for school. The target of evaluation at this time was a behavioral episode, “having breakfast with his/her family at home”. Indications of the end of an episode might be going to a different location, ending one activity and starting another, or a change in the people they are interacting with. The episodes the subjects are instructed to identify usually last between 15 min and two hours. Episode-based EMA is a similar concept to event-contingent EMA [2], but the subjects in episode-based EMA are asked to rate as many episodes as possible to cover almost the whole day as a continuous series of episodes, and the evaluated time span of each recording is from the start to the end of the most recent episode.

To eliminate individual subject differences when comparing the two EMA designs, the subjects were randomly assigned to one of the two EMA conditions and then crossed over to the other EMA condition. They were also apportioned evenly between weekdays and weekends. The first day of the week of the experiment was varied across subjects. After the time-based or episode-based assessments, the subjects visited my lab again to change the program of the watch-type computer from time-based to episode-based assessment, or vice versa. All the subjects performed a combination of both EMA assessments within a week.

Reconstructed fatigue and mood states: day reconstruction method

I used a recently proposed method, the DRM, in order to record reconstructed fatigue and mood states. The subjects filled out DRM questionnaires [9] about the days when EMA was conducted, once a day for three days. Regarding the day when episode-based EMA was conducted, the subjects were asked to complete the DRM questionnaires before going to sleep either on that day or the next day, and they were randomly assigned to the days. Regarding the days when time-based EMA was conducted, for the first day of time-based EMA they were asked to complete DRM questionnaires before going to sleep that day, and for the second day of time-based EMA, before going to sleep the next day. In this study, the DRM data were collapsed across the days on which the recordings were made because there were no systematic differences in the associations between DRM and EMA in the two reporting periods. The differences in the associations between DRM and EMA were investigated as ad hoc analyses by using the slopes of the multilevel models in which DRM was regressed on EMA. I could not find significant differences in the slopes between the two DRM conditions, except for depression as rated by DRM and by time-based EMA ($p = 0.040$), for which there were still significant associations between DRM and EMA for both DRM conditions (on that day, $p < 0.001$; the next day, $p < 0.001$).

2.2.3. Measurements

Both the EMA and DRM questionnaires consisted of items relating to behavior, and the intensity of fatigue and mood states. Subjects used questions on pull-down menus to select their behavior at each EMA recording, for example, what they were doing (i.e., activity), where they were (i.e., location), and who was interacting with them (i.e., companion). The categories of activity were “housework”, “eating”, “napping”, “entertainments”, “studying”, “exercising”, “relaxing”, “part-time job”, “shopping”, “commuting”, “taking a shower/bath”, and “other”. The categories of location consisted

of “home”, “school”, “transportation”, and “other” and those of companion were “alone”, “family”, “friends”, “significant other”, “colleague”, and “other”. The intensity of fatigue and mood states was assessed either at that time (time-based EMA) or in an episode (episode-based EMA). The time information of EMA recordings were automatically registered in the ED. In DRM recordings, the subjects were asked to recall a short diary of their day as a continuous series of behavioral episodes in the order in which they have experienced from waking up to going to bed by recording the situation information (i.e., activity, location, and companion), using the same categories as EMA recording, and the starting and ending time. They were not asked to recall the EMA recordings that they had made, but rather a sequence of behavioral episodes throughout the day. The intensity of fatigue and mood states were subsequently recalled in all the behavioral episodes they have first recorded. The intensity of fatigue in both EMA and DRM was assessed by simply asking the subjects to indicate this using a single question with the adjective “fatigued”. Mood states in both EMA and DRM were rated using the “Depression and Anxiety Mood Scale (DAMS)” [120], which was developed to measure anxious and depressive moods as separately as possible. The DAMS comprises nine adjectives representing mood states: “vigorous”, “gloomy”, “concerned”, “happy”, “unpleasant”, “anxious”, “cheerful”, “depressed”, and “worried”. From these items, anxious mood (sum of “concerned”, “anxious”, and “worried” scores), positive mood (sum of “vigorous”, “happy”, and “cheerful” scores), and negative mood (sum of “gloomy”, “unpleasant”, and “depressed” scores) were assessed. The depressive mood scores were obtained by combining the last two mood scores [(300 – positive mood) + negative mood]. These mood scores were rescaled in the range 0–100.

In both time-based and episode-based EMA assessments, symptoms (fatigue intensity and the DAMS items) were rated using a visual analog scale from 0 to 100 displayed on the screen of the ED. Each of the momentary symptoms was displayed one item at a time with the visual analog scale, and

the anchor words “none” and “most intense” were displayed at the respective ends of the scale. By manipulating the joystick of the ED, the subjects adjusted the length of the bar so that it corresponded to their subjective symptoms. Due to the limitation of display resolution, the 0–100 scale for the intensity of subjective symptoms increased at 5-point intervals. Reconstructed fatigue intensity and mood states in DRM assessments were recorded for each behavioral episode using a 100 mm visual analog scale anchored with the words “none” at the left end and “most intense” at the right end. An example of a subject’s fatigue intensity assessed by the two measurements is shown in Figure 2-1.

2.2.4. Data analysis

To investigate the correlations and differences between momentary and reconstructed assessments, EMA and DRM recordings were paired for each episode. Time-based EMA recordings were paired with DRM recordings if

- a. First, the subjects’ behavior (e.g., activity, location, and companion) recorded by time-based EMA corresponded to that recorded by DRM.
- b. Second, a time-based EMA recording point was within plus/minus 100% of the interval between the starting and ending times of a DRM recording.

For example, if a subject recorded by DRM having had lunch with his/her friend at school from 12:30 to 13:00, and if the same assessment by time-based EMA was registered between 12:00 and 13:30 (plus/minus 100% of the interval), then the two assessments recorded with DRM and time-based EMA were matched as one set of paired assessments. Episode-based EMA recordings were paired with DRM recordings for each episode if

- a. First, the subjects’ behavior (e.g., activity, location, and companion) recorded by episode-based EMA corresponded to that recorded by DRM.

- b. Second, the duration of an episode-based EMA recording overlapped with plus/minus 100% duration of a DRM recording.

Plus/minus 100% of the interval (duration) was chosen, from 25 to 200%, as the shortest allowance length to maximize the number of matched assessments.

2.2.5. Statistical analysis

The data set in the present study has a hierarchical structure, in which EMA and DRM recordings are nested within subjects. I therefore used multilevel modeling, which has been recommended for the analysis of data with a hierarchical structure [107], in order to investigate within-individual correspondence between EMA and DRM, using SAS Proc Mixed (SAS 9.1, SAS Institute Inc., Cary, NC). A p value less than 0.05 was considered significant.

Differences of time information between reconstructed recordings and momentary recordings

To investigate the differences of time information of two measurements, I first subtracted every starting time, ending time and time span of episode-based EMA recordings from matched DRM ones and then tested whether the differences are equal to zero or not using multilevel models with the differences as the dependent variable and with no predictor.

Correlations and differences between reconstructed recordings and momentary recordings

To test the correlations and differences between DRM and EMA, I used 2-level mixed MANOVA models, with DRM and EMA values as the dependent variables and type of assessment (i.e., DRM or EMA) as the predictor. The following four models were estimated for each symptom: model 1, which allows for unequal means and within-individual variability of DRM and EMA; model 2, which allows

for unequal within-individual variability but forces means to be equal; model 3, which allows for unequal means but forces within-individual variability to be equal; model 4, which forces both means and within-individual variability to be equal between DRM and EMA. Correlation coefficients were estimated and tested using model 1. Equality of means and within-individual variability of DRM and EMA were tested by comparing goodness of fit statistics of models 1–4. The change in -2 log-likelihood between the two nested models was tested using a chi-square test with degrees of freedom equal to the difference in the number of parameters. Equality of means was also tested by using estimates from model 1.

Effect of situational information on differences between reconstructed recordings and momentary recordings

To test whether or not reconstructed symptoms are systematically biased by certain situational information, I used multilevel models with the differences obtained by the subtraction of EMA recordings from matched DRM recordings as the dependent variable and situational information as the predictor either as a fixed or random effect and tested whether the effects of situational information on the differences are significant or not. The goodness of fit of the models was compared by calculating the difference in the -2 log-likelihood and testing it with a chi-square test with degrees of freedom equal to the difference in the number of parameters. A model with as few parameters as possible (i.e., a parsimonious model) was selected when the difference in -2 log-likelihood was not significant. I collected situational information about the subjects' activity, location, and companion for each episode, as described earlier. This information was converted to dichotomous variables (active/inactive behavior, outside home/at home, and alone/with someone). Activities were divided into active/inactive behavior according to whether the specific episode was sedentary behavior,

characterized by being in a sitting position without taking much exercise, or not [121].

2.3. Results

2.3.1. Subject characteristics

Thirty-five subjects were enrolled. Ten were excluded from further analyses because they could not complete their recordings due to a faulty recording device. Finally, 25 subjects (23 males and 2 females) were analyzed. The mean and standard deviation of age were 21.9 ± 2.4 years.

2.3.2. Recording profiles

The average compliance rate for time-based EMA was 90.97% ($SD = 8.34$), and the average number of recordings was 8.82 ($SD = 1.64$) per day. The average number of episodes per day recorded with episode-based EMA was 10.12 ($SD = 4.94$), and the average length of episodes was 83.72 min ($SD = 126.58$, median = 54, lower and upper quartile range = 17–106). The average number of episodes per day recorded by DRM when time-based EMA was conducted was 12.04 ($SD = 3.52$), and the mean episode length was 76.10 min ($SD = 77.73$, median = 60, lower and upper quartile range = 30–100). The average number of episodes per day recorded with DRM when episode-based EMA was conducted was 12.12 ($SD = 3.97$), and the mean episode length was 71.02 min ($SD = 79.85$, median = 45, lower and upper quartile range = 30–90). The number of DRM recordings in this study was similar to that in the previous study which presented DRM for the first time [the average number of episodes per day was 14.1 ($SD = 4.8$)] [9].

2.3.3. Matching profiles of reconstructed recordings and momentary recordings

Reconstructed behavioral episodes were matched with momentarily recorded behavioral episodes on the basis of subject's situational and time information for each episode. When the episode was matched on the basis of plus/minus 100% of the time information, $67.65 \pm 14.89\%$ of time-based

Table 2-1. Differences in time information between reconstructed recordings (DRM) and momentary recordings (episode-based EMA)

| | Differences between reconstructed DRM and episode-based EMA | | |
|---------------------|---|------------------|----------------|
| | Coefficient(<i>SE</i>) | <i>t</i> value | <i>p</i> value |
| Starting time (min) | 27.97(7.96) | $t_{24} = 3.51$ | 0.002 |
| Ending time (min) | 14.62(5.51) | $t_{24} = 2.65$ | 0.014 |
| Time span (min) | -15.53(7.46) | $t_{24} = -2.08$ | 0.048 |

DRM: day reconstruction method. EMA: ecological momentary assessment. *SE*: standard errors.

EMA recordings were detected in DRM recordings and $72.37 \pm 21.59\%$ of episode-based EMA recordings were found in DRM recordings. There was no statistically significant difference between these percentages, $t(24) = 0.97$, $p = 0.34$, suggesting that the behavioral episodes were remembered equally with DRM in both EMA conditions. Otherwise, when I used the time information strictly (i.e., based on plus/minus 0% of the time information) for matching, $45.61 \pm 15.62\%$ of time-based EMA recordings were detected in DRM recordings and $50.79 \pm 25.04\%$ of episode-based EMA recordings were found in DRM recordings. The number of matched episodes was significantly reduced when I changed the criteria for information about time from plus/minus 100% to a strict correspondence.

2.3.4. Differences of time information between reconstructed recordings and momentary recordings

I found that the starting and ending times of behavioral episodes recorded with DRM were significantly behind those of episode-based EMA (Table 2-1). The time span of episodes recorded with DRM was significantly shorter than that of episode-based EMA (Table 2-1).

Table 2-2. Correlations between reconstructed recordings (DRM) and momentary recordings (EMA)

| | Correlations between DRM and time-based EMA | | Correlations between DRM and episode-based EMA | |
|------------|--|----------------|---|----------------|
| | Coefficient | <i>p</i> value | Coefficient | <i>p</i> value |
| Fatigue | 0.31 | < 0.001 | 0.46 | < 0.001 |
| Depression | 0.52 | < 0.001 | 0.41 | < 0.001 |
| Anxiety | 0.31 | < 0.001 | 0.42 | < 0.001 |

DRM: day reconstruction method. EMA: ecological momentary assessment.

Table 2-3. Differences in mean levels and within-individual variability between reconstructed recordings (DRM) and momentary recordings (EMA)

| a) Differences between DRM and time-based EMA | | | | | | |
|--|-----------------------------|-------------|-----------------------------|------------------------------------|---------------|-----------------------------|
| | Estimated mean(<i>SE</i>) | | | Estimated variability(<i>SE</i>) | | |
| | DRM | EMA | <i>p</i> value ^a | DRM | EMA | <i>p</i> value ^b |
| Fatigue | 46.34(4.06) | 47.58(3.44) | 0.549 | 344.94(29.74) | 345.23(29.81) | 0.994 |
| Depression | 41.20(2.14) | 40.92(2.09) | 0.814 | 139.91(12.08) | 139.84(12.08) | 0.996 |
| Anxiety | 27.95(3.68) | 30.29(3.52) | 0.206 | 183.55(15.86) | 147.79(12.78) | 0.063 |
| b) Differences between DRM and episode-based EMA | | | | | | |
| | Estimated mean(<i>SE</i>) | | | Estimated variability(<i>SE</i>) | | |
| | DRM | EMA | <i>p</i> value ^a | DRM | EMA | <i>p</i> value ^b |
| Fatigue | 44.92(3.86) | 46.20(3.18) | 0.499 | 413.50(46.90) | 334.59(37.71) | 0.136 |
| Depression | 42.82(2.30) | 40.54(2.32) | 0.171 | 153.43(17.50) | 126.79(14.52) | 0.194 |
| Anxiety | 30.95(4.70) | 31.41(3.55) | 0.829 | 262.48(29.70) | 200.32(22.76) | 0.063 |

DRM: day reconstruction method. EMA: ecological momentary assessment. *SE*: standard errors. ^a*p* value for the mean difference. ^b*p* value for the change in $-2 \log$ -likelihood between the model that allows for unequal means and within-individual variability of DRM and EMA and the model that allows for unequal means but forces within-individual variability to be equal, reflecting the significance of within-individual variability difference.

2.3.5. Correlations and differences between reconstructed recordings and momentary recordings

There were significant correlations between reconstructed fatigue and momentary fatigue recorded with both time-based and episode-based EMA (Table 2-2). Significant correlations were also found between reconstructed and momentary mood states recorded with both time-based and episode-based EMA (Table 2-2). However, the levels of correlation between DRM and EMA recordings were low (Table 2-2).

The differences in the mean levels and within-individual variability between DRM and EMA recordings were not significantly different for either fatigue or mood states (Table 2-3). Equality of means was tested by using estimates from model 1 and there were no significant differences for all symptoms (Table 2-3).

2.3.6. Effect of situational information on differences between reconstructed recordings and momentary recordings

The multilevel model with situational information as a fixed effect was adopted on the basis of fit

Table 2-4. Effects of situational information on difference between reconstructed depression (DRM) and momentary depression (time-based EMA)

| | Difference between reconstructed depression and momentary depression | | |
|--|--|-------------------|----------------|
| | Coefficient(<i>SE</i>) | <i>F</i> value | <i>p</i> value |
| Intercept | -2.44(2.03) | | |
| Effect of situational information (activity) | | | |
| Active | Referent | | – |
| Inactive | 4.29(1.87) | $F_{1,20} = 5.26$ | 0.033 |

DRM: day reconstruction method. EMA: ecological momentary assessment. *SE*: standard errors.

statistics. There was a significant effect of activity (i.e., whether subjects' behavior was active or inactive; Table 2-4) on the differences between reconstructed and momentary (time-based EMA) depression. Reconstructed depression tended to be overstated compared to momentary depression when the subjects remembered their behavior was inactive. Neither the location (i.e., whether subjects were outside the home or at home) nor the companion (i.e., whether subjects were interacting with anyone or not) had any significant effect on depression recorded by time-based EMA. The effects of all the situational information on the differences between reconstructed and momentary (time-based EMA) fatigue and anxiety were not significant. There was no significant effect of situational information on the differences between reconstructed recordings and momentary recordings by episode-based EMA.

2.4. Discussion

In this study, correlations and differences in the mean and variability between DRM and EMA were systematically compared in the same subjects and for the same study period. In the previous studies [9,11,119], the comparisons of two measurements were examined based on averaged observations aggregated across a certain monitoring period because the observations rated with only time information could not be matched to each other at every single assessment. By acquiring simultaneous situation and time information for every episode, the two assessments could be paired at each observation and the details of changes over time in fatigue and mood states were investigated in this study.

The average number of recalled behavioral episodes per day recorded with DRM was not particularly different from that of the previous study, wherein the feasibility of DRM was shown using the same definition of behavioral episodes, although the total number of symptoms reconstructed by DRM was not exactly the same. Time information rated by DRM was significantly different from that by episode-based EMA and the percentage of matched episodes dropped significantly when I used strict time information without the plus/minus 100% allowance. These results suggest that the time information of episodes may not be accurately reconstructed by DRM conducted on that day or the next day, while we can relatively remember the situation information (e.g., activity, location, and companion) of episodes. DRM may not be a sufficient measurement when we search for accurate changes over time of fatigue and mood states related to a certain event (e.g., physiological information or behavioral episode). In the matching of DRM and EMA, I expected a higher matching rate between DRM and episode-based EMA because both of them were recorded by an episode, but the matching rate of DRM vs. episode-based EMA was not significantly higher than that of DRM vs. time-based EMA.

Statistical tests were made to address the correlations and the differences in the mean levels and within-individual variability of the two measurements. The results showed that there were significant correlations between DRM and EMA assessments and I could not detect statistically significant differences in mean and variability between DRM and EMA assessments. Nevertheless, I also found low levels of correlation between DRM and EMA. These results imply that we may be able to reconstruct the approximate intensities and overall changes of fatigue and mood states estimated over the entire study period on that day or on the next day, but detailed changes over time in fatigue and mood states cannot be remembered at the end of that day or the next day. Although DRM may be barely useful when researchers and clinicians would like to know the averaged levels or magnitude of variability of fatigue and mood states intensity over one or two days, it is not an appropriate substitute recording strategy to trace such detailed changes of fatigue and mood states. While there were low levels of correlation for fatigue and anxiety between DRM and time-based EMA, correlations between DRM and time-based EMA were higher for depression. Although the reason for this cannot be fully explained in the present study, I speculate that symptoms might be differently locked into specific episodes at the moment and reconstructing their intensities based on reconstructed behavioral episodes (i.e., by DRM) also varied across the symptoms.

In DRM measurements, subjects first describe their activities during the entire preceding day, and then they are asked to recall their experiences. Some studies indicate that reconstructing the context of daily activities is helpful in reducing recall bias [9,11,122]. However, the present study shows the magnitude of depression recorded by DRM to be affected by situational information when compared with that of time-based EMA as a standard. One possibility we can hypothesize for this result is that there may be systematic biases in DRM assessment and its biases may be related to certain situational information. Another possibility is that the biases may exist when we reconstruct

symptoms irrespective of whether the reconstruction is based on certain situational information or not. However, I was not able to find the effect of situational information in the case of the association between DRM and episode-based EMA despite the fact that both methods recorded symptoms related to situational information. This result may support the former possibility.

There are some limitations to this study. First, the subjects in this study were all healthy young people and included few females, so it is not possible to generalize the results for different populations or directly apply them in clinical settings. Although the within-individual variability of fatigue and mood states showed that they varied across the study period even in healthy adults (see also, e.g., Figure 2-1), changes of fatigue and mood states over time for healthy adults and the population in clinical settings would have different properties and this should be considered in further studies. Second, the correlation between momentary and reconstructed fatigue and mood states may be overestimated in this study, because the subjects rated their momentary experiences before making their retrospective ratings. However, we should note that changes over time of momentary fatigue and mood states could not accurately be reconstructed by DRM in spite of this advantage for recall. Finally, because a relatively small sample was available for analysis, this study had limited statistical power to detect differences between EMA and DRM, and the effect of situational information on reconstruction of symptoms in DRM.

In summary, the findings show that we are not able to reconstruct our diurnal time course (i.e., detailed changes over time) of fatigue and mood states with DRM, while the averaged mean and within-individual variability of momentary fatigue and mood states during the study period may be accessible with DRM. Also, I found that reconstructed depression intensity can be differently associated when the subjects are involved in active or inactive behavior. These results give us information about which method to select according to the study situation. This should also be taken

into consideration when evaluating fatigue and mood states in clinical settings. Further studies in clinical populations would clarify the nature of momentary and reconstructed self-reports for fatigue and mood states related to diseases.

PART 3: CO-VARIATION OF DEPRESSIVE MOOD AND LOCOMOTOR DYNAMICS

3.1. Introduction

Ecological momentary assessment (EMA) is a method of acquiring self-reported information about a person's subjective symptoms at the moment of recording and is used continuously in most cases to record instantaneous states of feeling without recall bias [2,3]. A paper-and-pencil diary was initially used for EMA, but it has been reported that the diary has the disadvantage of delayed data entry or even forward entry of data (i.e., faked compliance) [98]. Therefore, computerized EMA, which employs a computer as an electronic diary (ED), has been developed to avoid faked compliance; in this method, the input time is automatically registered in the device. Symptom diaries derived from computerized EMA are now generally regarded as the "gold standard" in the fields of psychiatric/psychosomatic medicine and has recently attracted increasing attention as an essential component for healthcare monitoring systems based on the information and communication technologies (ICTs) [86,123].

EMA scores for, e.g., depressive mood and fatigue, which are the key symptoms of depression [124] studied mainly in this study, inevitably fluctuate because of their repetitive nature. Characteristics of these fluctuations have been studied by examining the relationships among simultaneously measured self-reported symptoms [41,43,78]. Many studies have also investigated that the fluctuations in psychological states are associated with those in objective measures sampled momentarily. For example, psychological stress is reported to be concurrently associated with cardiovascular parameters from blood pressures [88,91,111] and salivary cortisol [28,29,30,31]. Also, the co-variations between pulmonary functions tested by a spirometer and positive/negative affect in

patients with asthma were reported [34]. In addition, some studies demonstrated that addictive behaviors such as smoking [47,48] and alcohol consumption [97,125] are related to the fluctuations in psychological states, e.g., positive/negative affect and craving. Furthermore, the associations of momentary psychological states with self-reported physical activity were demonstrated [45,46].

To my knowledge, however, whether EMA scores co-vary with other symptoms evaluated simultaneously and with locomotor activity (habitual physical activity) measured continuously and un-obstructively, which is an important hypothesis for the interpretation of fluctuations of EMA scores, has not been rigorously tested. Therefore, the primary aim of this study was to test this hypothesis for groups of healthy subjects in two ways: evaluating (null) convergent associations among self-reported mood and fatigue scores and concurrent associations of the self-reported scores with external criteria provided by objective measures of locomotor activity, as described below.

It is considered that momentary micro-fluctuations in behavioral data, specifically those in locomotor activity capturing bodily acceleration counts in a continuous fashion, reflect the dynamics of systems organizing human behavior and can be used to probe behavioral disorders, including mental illnesses [21,22,23,24,25,26,27]. Indeed, altered locomotor activity is one of the cardinal signs of psychiatric disorders and included in their diagnostic criteria [124]. For example, major depressive disorders are known to be characterized by the presence of symptoms associated with behavioral alterations, such as diminished physical activity, loss of energy, psychomotor retardation or agitation, and sleep disturbances [124]. In order to quantitatively evaluate such alterations, several actigraphic studies have been conducted in depression patients by showing the significant decrease in activity levels during daytime and disruption of the circadian rhythm [21,22,23]. Furthermore, it was recently found robust statistical properties of long-term (>7 days) locomotor activity altered in patients with major depressive disorder but not in healthy controls, reflecting increased intermittent bursts in the

activity counts, characterized by reduced activity levels associated with occasional bursts of locomotor activity [126,127]. Therefore, the quantitative and objective evaluation of the levels and/or the intermittency of locomotor activity could provide appropriate behavioral measures capable of probing alterations in patients' depressive mood and/or physical symptoms in daily life.

While this recent finding is thought to be important because it may lead to an objective evaluation of depressive disorders, whether changes in depressive mood are “momentarily” associated with such behavioral alterations probed by statistical properties of locomotor (or habitual physical) activity remains unknown. Thus, the secondary aim of this study was to search for the robust underlying associations between momentary depressive mood using EMA and local statistical properties of locomotor activity in healthy humans, regardless of group differences in age, lifestyle, and occupation by evaluating them with cross validation among various groups of subjects. Showing such associations is considered to be important because locomotor activity can be measured in a truly continuous manner, and finding objective correlates of momentary mood scores can compensate for a known drawback of EMA that the data based on sparse sampling may not provide contiguous data despite high levels of respondent load [9,11,117]. The continuous estimation of changes in depressive mood from locomotor activity might also help to detect the worsening of depressive mood even in healthy individuals at the beginning of pathogenic processes to depressive disorders.

3.2. Methods

3.2.1. Subjects

A total of 113 healthy subjects were enrolled in this study; 35 were adolescents, 44 were undergraduates, and 34 were office workers. The adolescents were junior high school students in the secondary school attached to the Faculty of Education of the University of Tokyo. The undergraduates were recruited from the Department of Physical and Health Education at the University of Tokyo. Full-time employees at the University of Tokyo were recruited as the office worker subjects. All subjects were given a thorough explanation of the purposes and potential risks of the study by well-trained researchers; then, the subjects signed an institutionally approved informed consent form. A written consent was also obtained from the adolescent subjects' parents. The study was approved by the research ethics committee of the Graduate School of Education, the University of Tokyo.

Of the 113 subjects, 28 were excluded from my analysis because their data were not measured properly owing to faults and/or subject error encountered while using the devices. I finally obtained data from 85 subjects consisting of 30 adolescents [4 males (M)/26 females (F); age, 13.6 ± 0.5 years], 31 undergraduates (25M/6F; 21.6 ± 2.3 years), and 24 office workers (24M/0F; 41.0 ± 9.2 years).

3.2.2. Materials, procedures, and measurements

Ecological momentary assessment

In this study, I used the EMA technique to examine momentary symptoms. The approach allows researchers to address subjects' behavior, psychological states, and physiological reactions at multiple time points as they are experienced in daily life. Collecting data in natural settings can enhance the validity of measurements, thus avoiding the pitfalls of retrospective recall, which highly distort self-report data collection [2]. In the present study, momentary symptoms were measured during the

study period using the EMA technique with the following procedures.

A watch-type computer (Ruputer, ECOLOG, 42 g; Seiko Instruments Inc., Tokyo, Japan) [39] was used as the ED to record momentary symptoms. This device has a 20 mm × 30 mm screen and a joystick and buttons as input devices (Figure 2-1A). The subjects were provided all appropriate instructions and also given manuals on how to use the device before the beginning of the study period; in addition, the subjects manipulated the device as a practice exercise until they became accustomed to it.

Adolescents: This group was instructed to complete the EMA questionnaires at randomly selected times within plus or minus 10 min of the predefined times (12:50, 16:30, and 20:00) over seven consecutive days. The subjects were prompted to complete the questionnaires by a beep signal. If they could not enter the recordings when the device beeped, they were allowed to postpone the recordings twice for 15 min each time. If the recordings were not performed within 30 min, the corresponding EMA questionnaires were canceled.

Undergraduates: The EMA questionnaires were recorded over two consecutive days. The subjects were prompted by a beep signal to complete the EMA questionnaires on average every two hours during their waking period. The signals were programmed to occur randomly within intervals ± 12 min of the predefined times (8:00, 10:00, 12:00, ...). The subjects were allowed to postpone the recordings twice for 10 min each time. If the recordings were not performed within 20 min, the corresponding questionnaires were canceled. The first experimental day of the week was randomly chosen, regardless of weekdays and weekends, so that the study periods can cover all possible day-conditions.

Office workers: EMA questionnaires were recorded on average every four hours during their waking period, over seven consecutive days. The signals were programmed to occur randomly within

intervals ± 24 min of the predefined times (8:00, 12:00, 16:00, ...). The subjects were allowed to postpone the recordings twice for 10 min each time. If the recordings were not made within 20 min, they were canceled.

In addition, all subjects were asked to record the EMA questionnaires when they woke up and before they went to bed by choosing “waking up” or “going to bed”, from the menu. The selection of the latter item suspended the prompt for recordings until the former item was selected so that to the subjects’ sleep is not disturbed. Furthermore, they were instructed to remove the device before bathing, showering, or any other activity likely to damage the device. When they removed or wore the device, they were requested to select “taking off” or “putting on”, respectively, from the menu as well as the reason for removing the device. The adolescents were asked to conduct EMA recordings (i.e., event-based EMA [2]) in addition to the scheduled EMA (i.e., time-based EMA [2]) when they felt bad or had physical symptoms, such as severe fatigue, headache, or stomachache. Note that these data were analyzed without distinguishing between the two types of EMA because I obtained the consistent results whether such data points were included or not.

The EMA questionnaires assessed fatigue and mood states. The recording times were automatically registered in the device. Fatigue intensity and mood states were rated using a visual analog scale from 0 to 100, which was displayed on the device screen. The EMA questionnaires of momentary symptoms were randomly ordered, and each of them was displayed one at a time using a visual analog scale; the anchor words “none” and “most intense” were displayed at the respective ends of the scale. The subjects adjusted the length of the scale by manipulating the joystick of the device so that it corresponded to their subjective symptoms. Owing to the limited display resolution, the 0–100 scale for the intensity of subjective symptoms increased at 5-point intervals. Fatigue intensity was assessed by instructing the subjects to score their current fatigue level by presenting a single question

with the adjective “fatigued”. Mood states were rated using the Depression and Anxiety Mood Scale (DAMS) [120], which was developed to measure anxious and depressive moods as separately as possible. The DAMS comprises the following nine adjectives representing mood states: “vigorous”, “gloomy”, “concerned”, “happy”, “unpleasant”, “anxious”, “cheerful”, “depressed”, and “worried”. These nine items were used to calculate anxious and depressive mood scores (see 2.3.3 for details).

Assessment of locomotor activity

This watch-type device is also equipped with an activity monitor, which is analogous in performance to the commercial actigraph (Ambulatory Monitors Inc., Ardsley, NY, USA), widely used in clinical fields [21,126,127,128]. The sensor for assessing locomotor activity is a uni-axial piezo-electric accelerometer capable of detecting small changes in wrist acceleration (≥ 0.01 G/rad/s), which makes it possible to register even slight movements. All subjects were instructed to wear this device on the wrist of their respective non-dominant hand throughout the study period, except while bathing or performing rigorous exercises. Zero-crossing counts were accumulated for every 1-min epoch for the undergraduates and office workers. For adolescents, acceleration counts were accumulated for every 10 seconds; these activity data were aggregated to make 1-min values to ensure compatibility with the other groups. Locomotor activity data for periods in which the subjects were not wearing the device or sleeping were excluded from the analysis. An example of locomotor activity data is shown in Figure 3-1.

3.2.3. Data analysis

Local statistics of locomotor activity

I calculated the local statistics of locomotor activity data up to the third-order moment, but used only

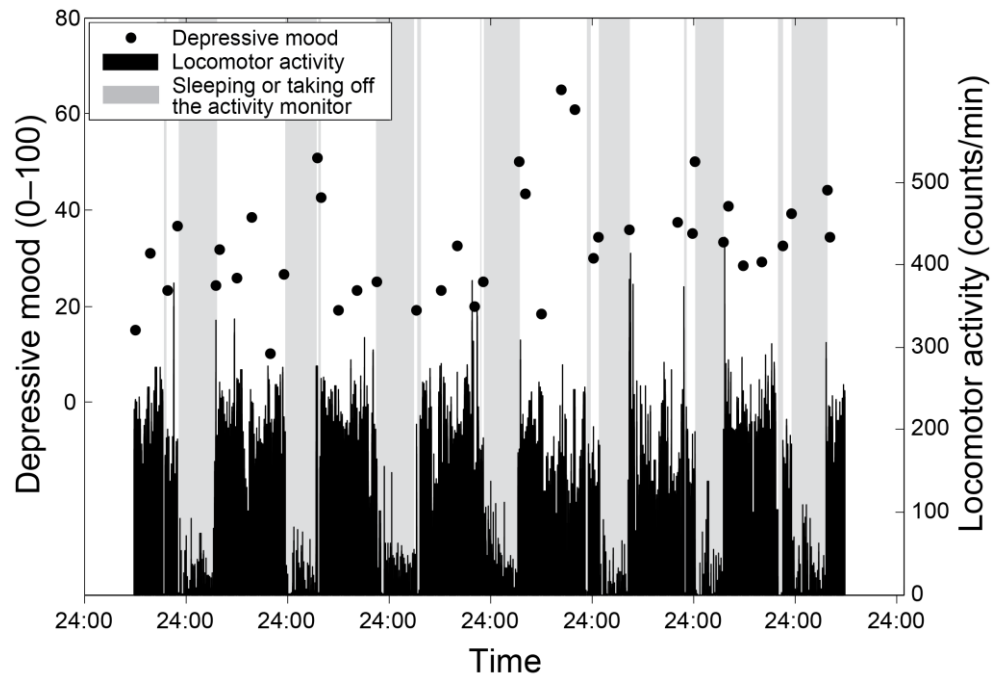


Figure 3-1. Fluctuation of depressive mood and locomotor activity in an office worker. Filled circles indicate depressive mood scores (y-axis on the left side) recorded by ecological momentary assessment (EMA) over seven consecutive days. The locomotor activity measured simultaneously (zero-crossing counts within every 1 min time interval) is also shown (y-axis on the right side). The periods shaded in gray are times during which the subject was sleeping or had taken off the device.

the mean and skewness because the first and third-order moments are considered to be sufficient to characterize the observed data. While the standard deviation (i.e., the second-order moment) is a standard measure characterizing variability of given data, it is not appropriate for the case where the data do not obey a normal distribution; the distribution of locomotor activity has nonnegative values, leading to a positively-skewed distribution. Indeed, the standard deviation did not play major roles in predicting depressive mood scores when it was included in the model [see Figure S1 in supplementary information (SI)]. In contrast, the skewness, as a measure of asymmetry of a distribution, is thought to be more appropriate to characterize the observed asymmetry. Lower or higher mean activity levels could quantify the states related to psychomotor retardation or agitation, respectively [21,23]. Higher skewness could characterize the intermittent bursts of locomotor activity [126,127,129].

In addition, I calculated the skewness from the detrended locomotor activity data, where a trend of activity data is subtracted by fitting the first-order polynomial (i.e., linear) function before calculation (see Text S1 and Figure S2 for the detailed procedures in SI). This detrending was aimed at eliminating effects of non-stationarity due to, e.g., daily activities; I examined the effects up the third-order polynomials without finding systematically different results (see Figure S1 in SI). Thus, I obtained three local statistics of locomotor activity, the mean, skewness, and (linearly) detrended skewness for evaluating the associations with the scores of EMA recordings.

Choice of size and location of time frame

The local statistics were calculated from the neighboring data of each EMA within a certain time frame with size N and location K . The value of K (min) represents the central position (location) of the time frame against the timing of EMA recordings, and the size N (min) indicates the data points used to calculate the statistics. Thus, the combination of N and K indicates the time frame starting from $N/2 - K$ points (min) prior to each EMA recording and ending at $N/2 + K$ points (min) after the recording. Therefore, the case where $N = 60$ and $K = -30$ indicates the time frame starting from 60 min prior to an EMA recording to the timing of the recording.

The choice of the size of the time frame may be important because they could have a significant impact on the robustness of the statistics and their temporal coincidence with the symptoms. Theoretically, the larger the size of the time frame, the greater is the stability and reliability of the estimates. However, the choice of larger time frames with less localized temporal positions may obscure the information on the momentary associations between locomotor activity and self-reported symptoms. In addition, the choice of the location plays an important role to investigate the casual association, whether locomotor activity precedes or follows changes in momentary symptoms. Thus,

careful consideration of such a trade-off is required, while the optimal choice of combination of the size and location is in advance unknown. Therefore, I varied the size N and location K of the time frame to examine their effects.

3.2.4. Statistical analysis

The data set in the present study has a hierarchical structure, in which EMA recordings were conducted at multiple times during the study period; thus, the recordings and the corresponding local statistics of locomotor activity were nested within the subjects. Therefore, I used multilevel modeling, which is an extension of traditional regression models and has been recommended for the analysis of data with a hierarchical structure including the within- and between-individual levels (e.g., individual differences such as age, gender and occupation) [107], using SAS Proc Mixed (SAS 9.2, SAS Institute Inc., Cary, NC). Multilevel modeling can also handle unbalanced data in which the number of the EMA recordings is different across subjects and can include both random and fixed effects together in the same model. A p value less than 0.05 was considered significant. For the univariate multilevel modeling, a false discovery rate (FDR) procedure [130] was performed at a q value of 0.05 to correct for the multiple comparisons. All models were estimated from the data set sampled from the three groups.

Convergent associations among symptoms

To evaluate the convergent associations among self-reported moods and fatigue, simultaneously measured by EMA, I calculated correlations among their scores using 2-level mixed MANOVA models with EMA scores as the dependent variables and the type of symptoms (i.e., depressive mood, anxious mood, and fatigue) as the predictor. Various studies demonstrate the co-variation between

depressive and anxious moods, suggesting the possibility to categorize them into the same class of a broader negative affect dimension [131,132,133]. Indeed, the validation study of DAMS confirms the significant positive correlation between these two mood scores [120]. Fatigue is known to have an extremely high comorbidity rate with depressive as well as anxious moods. For example, fatigue is a common complaint in depressive patients [134,135,136,137]. In chronic fatigue syndrome, the high rate of its comorbidity with both moods has been reported [138]. Co-variations of fatigue with these moods have also been reported in patients with cancer [139]. Therefore, these three symptoms would be expected to co-vary with each other if the subjects properly performed the EMA recordings; otherwise, such correlations should be absent.

Concurrent associations of symptoms with local statistics of locomotor activity

To investigate the concurrent associations between self-reported symptoms and simultaneous locomotor activity data, the following multilevel models were estimated. In all models, the scores of EMA recordings were treated as the dependent variables, and the corresponding local statistics of locomotor activity were modeled as the predictors. Firstly, I estimated univariate multilevel models to test the simple correlations between the symptoms and local statistics. In addition, I also examined the effects of the choice of time frame to find an optimal combination of size and location for evaluating the local statistics. With regard to the size of the time frame, I considered 12 different conditions: $N = 120, 110, 100, \dots, 20, 10$ min. I also considered 25 different locations: $K = -60, -55, -50, \dots, 55, 60$ min. I considered a total of 300 combinations ($12 \text{ sizes} \times 25 \text{ locations}$) for each symptom and statistic. The FDR with the q value of 0.05 was used as the multiple comparison adjustment.

Secondly, all possible multivariate multilevel models were evaluated by considering the three predictors and their interaction terms. However, the terms consisting of both skewness and detrended

skewness were not considered; the interaction terms I considered were: mean \times skewness and mean \times detrended skewness. Furthermore, I considered both fixed (i.e., the slope of the predictor does not vary across individuals) and random (i.e., the slope varies across individual) effects for all predictors. In this part, I focus on the results for depressive mood because I could not find a significant association of locomotor activity with fatigue or anxious mood.

I adopted the following procedures to evaluate the goodness of fit of the models. If the inclusion of a random effect into a predictor significantly improved the residual variance of the data explained by the model, I considered the random effect for that term [7]. After determining the type of effect of all predictors for all models, I compared the goodness of fit by using two types of tests depending on the structure of the models. While comparing the models with nested or inclusive relations, where all predictors used in one model were a subset of the predictors used in another model, I tested the deviance statistic [105]. This approach tests the difference in the -2 log-likelihood on the basis of the chi-square test with degrees of freedom equal to the difference in the number of model parameters to be estimated. If the difference in the -2 log-likelihood was not significant, the model with fewer parameters was selected as the preferable one on the basis of the *principle of parsimony*. For the non-nested models, I used a statistical test based on the Akaike information criteria (AIC), in which the difference of AIC between models is evaluated using the z -test [140].

To determine the best-fitting model for describing the variations of depressive mood scores, I used a time frame of 60 min centered around the EMA recording (i.e., $N = 60$, $K = 0$) on the basis of the investigation of univariate multilevel models (see Results). After determining the best-fitting model, I also considered the effects of size and location by varying them. I considered 12 different sizes ($N = 120, 110, 100, \dots, 20, 10$) keeping the location of the frame ($K = 0$). I also considered 13 different locations ($K = -30, -25, -20, \dots, 25, 30$) with the size set at 60 min.

Cross validation for association of depressive mood with local statistics of locomotor activity

To cross validate the associations of depressive mood scores with local statistics of locomotor activity among the three groups, an ad hoc analysis was performed for the best-fitting model. I added a categorical variable representing the groups (adolescents, undergraduates, and office workers) to the best-fitting model in order to test whether the concurrent associations differed among categories.

3.3. Results

3.3.1. Recording profiles

The mean compliance rate, which is defined as the percentage of the number of completed recordings within 30 min from the first beep signal against the total number of time-based EMA during the study period, was 94.54% ($SD = 5.96$), and the mean number of recordings was 5.40 ($SD = 1.58$) per day for the adolescents [the mean number of event-based EMA recordings was 1.19 ($SD = 0.88$) per day for this group]; for the undergraduates, they were 92.03% ($SD = 8.57$) and 8.98 ($SD = 1.43$), respectively, and for the office workers, they were 95.40% ($SD = 4.54$) and 6.05 ($SD = 4.49$), respectively. The compliance rate was considerably high for all groups without significant differences (ANOVA, $F_{2,82} = 1.92$, $p = 0.153$). The mean scores of the three symptoms and the three local statistics of locomotor activity are summarized in Table 3-1.

Table 3-1. Statistics of self-reported symptoms and local statistics of locomotor activity

| | Mean(SD) | | | | Multiple comparison |
|-------------------------------|-----------------|------------------|---------------------|---------------------|---------------------|
| | All | Adolescents (G1) | Undergraduates (G2) | Office workers (G3) | |
| Self-reported symptoms | | | | | |
| Depressive mood | 39.93(10.57) | 40.00(12.09) | 40.31(9.62) | 39.34(10.13) | |
| Anxious mood | 31.54(19.25) | 33.81(21.67) | 31.40(18.09) | 28.88(17.90) | |
| Fatigue | 44.47(18.43) | 45.54(19.04) | 47.42(18.18) | 39.33(17.65) | |
| Local statistics of LA | | | | | |
| Mean | 144.44(22.31)** | 152.76(17.97) | 149.59(22.00) | 127.38(18.72) | G3 < G1, G2 |
| Skewness | 0.04(0.27)** | -0.11(0.22) | 0.11(0.28) | 0.14(0.22) | G1 < G2, G3 |
| Detrended skewness | 0.06(0.23)* | -0.03(0.20) | 0.08(0.25) | 0.14(0.21) | G1 < G3 |

LA: locomotor activity, SD : standard deviation. ** and * indicate significant group difference at $p < 0.01$ and $p < 0.05$, respectively. Note that all local statistics of LA were evaluated using a 60-min time frame centered around ecological momentary assessment recordings.

Table 3-2. Convergent associations (correlations) among self-reported mood and fatigue scores

| | depressive vs. anxious mood | depressive mood vs. fatigue | anxious mood vs. fatigue |
|----------------|--------------------------------|--------------------------------|-----------------------------|
| All | 0.39* | 0.26* | 0.18* |
| Adolescents | 0.32* | 0.25* | 0.17* |
| Undergraduates | 0.39* | 0.38* | 0.27* |
| Office workers | 0.60* | 0.25* | 0.19* |

* $p < 0.01$ **3.3.2. Convergent associations among symptoms**

Depressive mood showed significant positive correlation with fatigue ($r = 0.26$, $p < 0.001$) and anxious mood ($r = 0.39$, $p < 0.001$) for the data sets sampled from all groups. The correlation between fatigue and anxious mood was also positive and significant ($r = 0.18$, $p < 0.001$). These relations were consistent across groups (Table 3-2), indicating that all subjects properly performed the EMA recordings. The results also suggest the convergent validity of self-reported symptoms by EMA.

3.3.3. Concurrent associations of depressive mood with local statistics of locomotor activity*Univariate multilevel modeling*

The univariate multilevel analysis revealed simple correlations between the depressive mood scores and the local statistics (Figure 3-2). The mean and skewness as well as the detrended skewness showed significant correlations (FDR, $q = 0.05$). However, such relations depended significantly on the size and location of the time frame, especially for skewness (Figure 3-2). The mean levels of locomotor activity were negatively associated with depressive mood over a wide range of locations (60 min before to 55 min after) around the EMA recordings, and the associations were not significantly affected by the size of the time frame (data length 120–10 points; Figure 3-2A). This

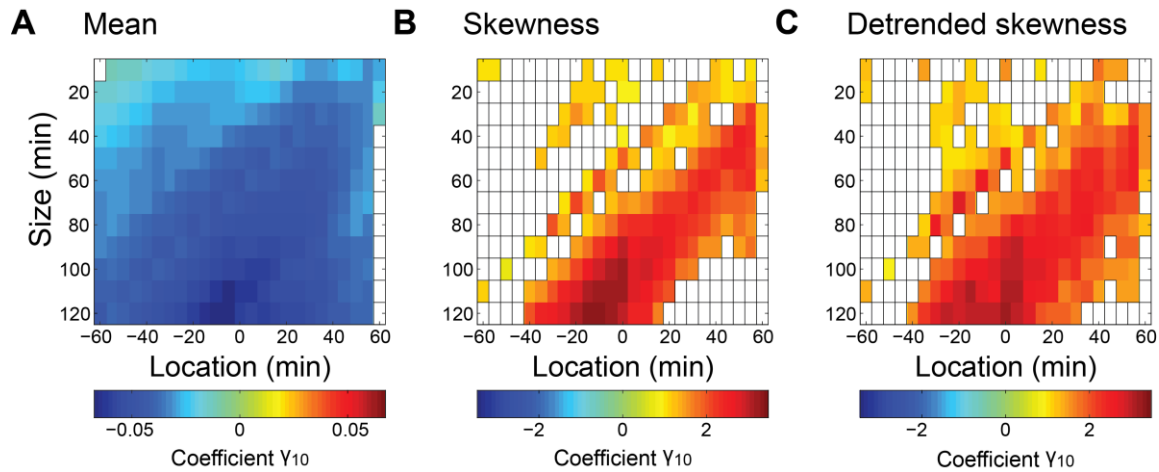


Figure 3-2 Dependency of univariate model coefficients for depressive mood on different time frames. The estimated values of the univariate model coefficient for depressive mood scores are shown in a colored matrix form consisting of 25 rows (different locations) and 12 columns (different sizes). Each grid cell indicates some specific location and size of a time frame used for calculating the local statistics of locomotor activity. A color in each cell represents the value of the model coefficient (γ_{10}) of the predictors: (A) Mean, (B) Skewness, and (C) Detrended skewness. False discovery rate with a q value of 0.05 was used as the multiple comparison adjustment. Only significant cases are shown by colors. Color bars indicate the value of the model coefficient. Note that the univariate model used for the analysis is as follows; Depressive mood score $_{ij} = \gamma_{00} + \gamma_{10}$ (Local statistics of locomotor activity $_{ij}$) + $\zeta_{0i} + \varepsilon_{ij}$ (see Results for details).

indicates that activity levels tend to be low with a higher level of depressive mood, and this relationship is consistent and robust over a long period. On the other hand, skewness (also detrended skewness) exhibited positive correlations in a limited range centered around the EMA recordings (e.g., from location -15 to 55 min with size 60 points for skewness, and -25 to 60 min for detrended skewness; Figure 3-2B, C). This implies that depressive mood is associated with intermittent patterns of locomotor activity even in healthy populations, and this relationship appears instantaneously with the worsening of depressive mood. However, such a relationship disappeared with a small size of the time frame, which is partly attributed to the lack of statistical stability of skewness. Considering these results, I used the time frame centered around the EMA recordings with a size of 60 min while constructing the best-fitting model for depressive mood and examined the effects of the choice of time frame. For the univariate modeling, the coefficient of each predictor was fixed across the subjects

because the inclusion of the random effect did not have a large effect on the results.

Determination of best-fitting model

Of all multilevel models evaluated, a model consisting of local mean and detrended skewness as the predictors was selected as the best-fitting model to describe the variations of depressive mood scores,

Table 3-3. Associations of depressive mood with local statistics of locomotor activity

| | Coefficient(SE) | F value | p value |
|---|-----------------|---------------------|---------|
| <i>All data set</i> | | | |
| Intercept (γ_{00}) | 45.74(2.40) | | <0.001 |
| Coefficient of the local mean (γ_{10}) | -0.04(0.01) | $F_{1,1386} = 9.35$ | 0.002 |
| Coefficient of the local detrended skewness (γ_{20}) | 1.40(0.69) | $F_{1,1386} = 4.18$ | 0.041 |
| <i>Cross validation across the three groups</i> | | | |
| Intercept (γ_{00}) | | $F_{1,82} = 0.08$ | 0.920 |
| Adolescents | 47.28(3.95) | | <0.001 |
| Undergraduates | 45.25(4.30) | | <0.001 |
| Office workers | 45.24(4.20) | | <0.001 |
| Coefficient of the local mean (γ_{10}) | | $F_{1,1382} = 0.07$ | 0.935 |
| Adolescents | -0.05(0.02) | | 0.028 |
| Undergraduates | -0.04(0.02) | | 0.140 |
| Office workers | -0.04(0.02) | | 0.066 |
| Coefficient of the local detrended skewness (γ_{20}) | | $F_{1,1382} = 2.88$ | 0.056 |
| Adolescents | 2.49(0.99) | | 0.012 |
| Undergraduates | 2.26(1.42) | | 0.110 |
| Office workers | -1.24(1.29) | | 0.336 |

SE: standard errors. The best-fitting model is as follows: Depressive mood scores_{ij} = $\gamma_{00} + \gamma_{10} (\text{Mean}_{ij}) + \gamma_{20} (\text{Detrended skewness}_{ij}) + \zeta_{0i} + \zeta_{1i} (\text{Mean}_{ij}) + \varepsilon_{ij}$. For cross validation, I assumed the coefficients (i.e., slopes) γ_{10} and γ_{20} are different across groups.

where the predictor “mean” and intercept had a random effect. The best-fitting model is as follows: Depressive mood scores_{ij} = $\gamma_{00} + \gamma_{10} (\text{Mean}_{ij}) + \gamma_{20} (\text{Detrended skewness}_{ij}) + \zeta_{0i} + \zeta_{1i} (\text{Mean}_{ij}) + \varepsilon_{ij}$, where the Depressive mood scores_{ij} indicates a score of depressive mood at the *j*th recording for the *i*th subject; Mean_{ij} and Detrended skewness_{ij} are the local mean and detrended skewness of locomotor activity corresponding to the *j*th EMA recording (depressive mood score) for the *i*th subject, respectively; γ_{00} is the average intercept across all subjects; γ_{10} and γ_{20} are the average regression slopes across all subjects for each predictor; the random terms, ζ_{0i} and ζ_{1i} , are the between-individual residuals; and ε_{ij} is the within-individual residual. As expected from the results of univariate multilevel analysis, the coefficient of both local statistics was significant (Table 3-3), with a negative value (-0.04 ± 0.01 , $p = 0.002$) for the mean and a positive value (1.40 ± 0.69 , $p = 0.041$) for the detrended skewness.

Although the best-fitting model marked the minimum AIC value, a model incorporating local mean with a random effect ($F_{1,1386} = 9.80$, $p = 0.002$) and skewness with a fixed effect ($F_{1,1386} = 3.20$,

Table 3-4. Statistics of goodness of fit of best-fitting model and equally well-fitting models

| | Coefficient(SE) | F value | p value | AIC | p value* |
|---|-----------------|---------------------|---------|---------|----------|
| Best-fitting model | | | | 11960.8 | |
| Intercept | 45.74(2.40) | | <0.001 | | |
| Coefficient of local mean | -0.04(0.01) | $F_{1,1386} = 9.35$ | 0.002 | | |
| Coefficient of local detrended skewness | 1.40(0.69) | $F_{1,1386} = 4.18$ | 0.041 | | |
| Well-fitting model | | | | 11961.8 | 0.856 |
| Intercept | 45.92(2.40) | | <0.001 | | |
| Coefficient of local mean | -0.04(0.01) | $F_{1,1386} = 9.80$ | 0.002 | | |
| Coefficient of local skewness | 1.21(0.68) | $F_{1,1386} = 3.20$ | 0.074 | | |

SE: standard errors. AIC: Akaike information criteria. * p value for the change in AIC between the best-fitting model and each equally well-fitting model evaluated using the z-test.

$p = 0.074$) was equally well-fitted to the data without any significant difference from the best-fitting model (see Table 3-4 for details). The common feature represented in these models is the intermittency of locomotor activity, characterized by a combination of reduced activity with occasional bursts of locomotor activity.

Dependency on size and location

I demonstrate the robustness of the best-fitting model with regard to the choice of the size and location of the time frame by showing the dependency of the model coefficients on this choice (Figure 3-3). The coefficient of the “mean” was consistently significant with negative values (between -0.06 and -0.04) over a broader range (from location -30 to 20 min with size 60 points; Figure 3-3A). Furthermore, this association was also robust against variation in sizes ($120-10$ points; Figure 3-3B). In the case of “detrended skewness”, the significant positive coefficients were localized in a limited area around the EMA recordings (from 0 to 30 min with 60 points; Figure 3-3C), and this association was confirmed only at a larger size of the frame (data length > 50 points; Figure 3-3D).

3.3.4. Cross validation for association of depressive mood with local statistics of locomotor activity

To examine the cross validation for the associations of depressive mood scores with local statistics of locomotor activity across the three groups, I considered one categorical variable representing “group”. There was no significant difference in the coefficient for either predictor (Table 3-3) among the groups, which supports the cross validity of the results. However, careful investigation of this model demonstrated that the inclusion of the categorical variable and the separate evaluation of the

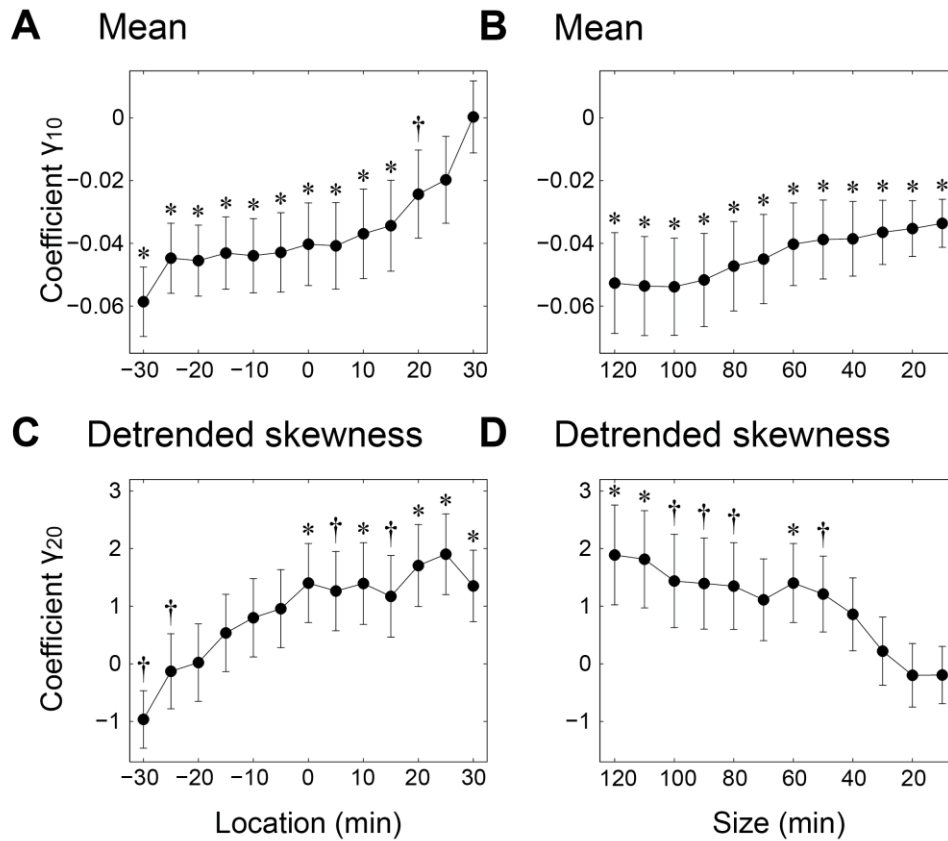


Figure 3-3. Dependency of best-fitting model coefficients for depressive mood on different time frames. The estimate of the model coefficient of the predictor: (A) Mean (γ_{10}) and (C) Detrended skewness (γ_{20}) of the best-fitting model for depressive mood scores ($\text{Depressive mood scores}_{ij} = \gamma_{00} + \gamma_{10} (\text{Mean}_{ij}) + \gamma_{20} (\text{Detrended skewness}_{ij}) + \zeta_{0i} + \zeta_{1i} (\text{Mean}_{ij}) + \varepsilon_{ij}$) as a function of the location. In panels (A) and (C), the size of the time frame was fixed at 60 min. The same is shown in panels (B) and (D), except for the dependency on the size of the time frame. For evaluation, the location of the time frame was fixed at 0 min; thus, the frame was centered around the ecological momentary assessment recordings. Error bars indicate standard error of the coefficients. The asterisk and dagger indicate significant cases at the 0.05 and 0.10, respectively.

best-fitting model for each group led to insignificant model coefficients, except for the adolescents.

This is caused by the difference in the sample size.

3.4. Discussion

In this study, I investigated the nature of within-individual temporal variations in self-reported symptoms assessed using the EMA technique. EMA is known to be a powerful tool to overcome retrospective biases existing in recall-based self-reports and thus has recently attracted massive attention [2,3]. Many studies have demonstrated the strength of EMA in terms of the minimization of recall biases [108,109,110] and the correlation with other objective measures [28,29,30,31,34,88,91,111]. However, the concurrent associations of self-reported symptoms by EMA with objective real-time measures were not rigorously examined. Therefore, I investigated such associations between self-reported symptoms, especially depressive mood, and locomotor activity data, which can be measured in a truly continuous fashion. I then found the robust (against differences in groups of healthy subjects and the methods of EMA) within-individual temporal associations between fluctuations in depressive mood in daily life and the patterns of locomotor activity. I believe this finding could provide a novel insight into the psychophysiological correlates of “variations” of EMA recordings and a possibility to develop a continuous monitor for changes in depressive mood from locomotor activity, as discussed below.

I confirmed the concurrent association of depressive mood with alterations in locomotor activity and the convergent associations among self-reported symptoms recorded simultaneously. In addition, I demonstrated the consistency and robustness of such associations across populations with different age groups, lifestyles, occupations, and so on and proved the existence of cross validity. These verifications provide strong evidence that the within-individual temporal variations in self-reported symptoms measured by EMA are not merely random but may partially originate from and reflect the alterations in psychophysiological states in daily life.

My systematic approach to modeling the temporal variations in depressive mood scores using the

local statistics of locomotor activity for the first time revealed the “momentary” associations of the depressive mood with the levels and patterns of locomotor activity. The intermittent bursts of locomotor activity were important for describing depressive mood even in healthy subjects, which is consistent with previous studies on major depressive disorders [126,127]. Considering the existence of the cross validity in these associations, I suggest that the present relations would generally hold for a wider range of populations. Whether this could be further extrapolated to patients suffering from depressive disorders would be the next important challenge.

The associations between depressive mood and alterations in the levels and patterns of locomotor activity have been shown in epidemiological samples [141,142,143,144,145] and clinical depression [126,127], but the underlying mechanism is unknown. Although the reasons for the covariant associations between depressive mood scores and behavioral alternations are still unclear, the present analyses of data with high temporal resolution clarify the mechanism, at least partially. As shown in Figure 3-2, there is less or virtually no association in the case of skewness measures reflecting intermittent activity patterns (shown in the upper left corner of the size vs. location plots). This implies that the association of the higher order statistics of locomotor activity data, even when recorded just before the EMA recordings, with the subsequent depressive mood scores is weak. Thus, it is most unlikely that behavioral alteration precedes momentary mood change, which is supported by higher coefficients for the detrended skewness of the best-fitting model at positive or posterior temporal locations (Figure 3-3C). The temporal characteristics of the covariant associations indicate that the behavioral alteration begins to be concurrent with mood change. This concurrency suggests that in addition to a possible effect of variations in depressive mood (or more instantaneous *feeling*) on the subsequent behavior, there is a more direct neurophysiological link between the changes in body state and (depressive) feeling, as proposed by Damasio [146]. On the other hand, the mean

activity levels were associated with depressive mood over relatively longer time scales.

Revealing the psychophysiological origin of the temporal variations in self-reported symptoms in terms of behavioral alterations is considered to be highly significant. EMA is known to have the critical drawback of sparse sampling over time; the number of data points recorded in a day is limited. In contrast, locomotor activity can be measured in a truly continuous fashion using a simple device. Therefore, the findings concerning the existence of psychobehavioral correlates between EMA recordings and behavioral alterations might be able to compensate for this drawback and contribute to the continuous estimation of changes in depressive mood in daily life, leading to the early detection or novel treatment of depressive disorders.

Several limitations should be noted. The populations in this study were limited to healthy subjects with different gender and age distributions. Another limitation is the inconsistency in the EMA protocols among groups leading to differences in the number of recordings and the time of day. Therefore, careful consideration is needed for generalizing the findings of this study to other populations, especially patients with psychiatric disorders such as depression. Also, the model coefficients of the local statistics for the depressive mood scores were not consistently significant in all groups, which is considered to be caused by the small sample size. Therefore, further studies with a large number of subjects and data points would be necessary to rigorously examine cross validity. Finally, other local statistics of locomotor activity that can capture the intermittency in locomotor activity more robustly, such as entropy-type nonlinear statistics, should be considered.

In conclusion, this is the first study providing an insight into the psychophysiological origin of within-individual temporal variations in momentary symptoms assessed by EMA by evaluating co-variant properties with continuous locomotor activity as the external criteria. I confirmed the robust concurrent associations between depressive mood and behavioral alterations and the

convergent associations among self-reported symptoms. These findings indicate that temporal variations in momentary symptoms are not random but reflect changes in the underlying psychophysiological variables, represented as changes in the statistical properties of locomotor activity. I believe that the finding on the concurrent changes in depressive mood and locomotor activity may contribute to the continuous estimation of changes in depressive mood and early detection of depressive disorders.

PART 4: CONTINUOUS ESTIMATION OF DEPRESSIVE MOOD FROM LOCOMOTOR DYNAMICS

4.1. Introduction

Major depressive disorder (MDD) is a severe psychiatric disorder that is characterized by the presence of mood disturbances (either depressed mood or a loss of interest or pleasure in daily activities) consistently for more than several weeks [124]. In addition to these mood disturbances, behavioral alterations, including diminished activity, loss of energy, and psychomotor retardation or agitation, are remarkable co-occurring symptoms [124]. In fact, many epidemiological studies of depressive disorders that used traditional paper-and-pencil self-report assessments have demonstrated the presence of altered physical activity levels [141,142,144,145,147] as well as their significant associations with the severity of the diseases [143,148]. As a more quantitative approach, many studies on MDD have collected continuous measurements of physical activities in daily life using an activity monitor and demonstrated a variety of behavioral alterations, including lower activity levels during daytime [21,23,149,150], sleep disturbances [150,151], and disruption of the circadian rhythm [21,22,152] as well as their improvement over the course of clinical treatment [21,23,153]. However, these studies mainly focused only on the alterations of physical activity levels or their rhythmicity, and the complete details of dynamical properties, which may contain richer information regarding pathological states, have not been examined.

In this context, recently, the so-called locomotor activity (>1 week), i.e., the spontaneous physical activity in daily life, in patients with MDD was measured and it was reported that patients with depression exhibited more intermittent behavioral patterns that were characterized by reduced mean activity levels associated with occasional bursts of locomotor activity compared with healthy

subjects [126,127]. These findings suggest that the statistical properties of intermittent locomotor dynamics can be important and useful objective markers of MDD, and there is a novel possibility of continuously monitoring the pathological states of this disorder based on behavioral dynamics. However, although these studies were successful in providing a biobehavioral measure for MDD based on long-term locomotor activity data, this is not sufficient for continuous monitoring because of the lack of temporal resolution and correlation with symptoms (e.g., subjective depressive mood). To capture rapid changes in pathological states in much shorter time frames (e.g., daily or within-day scales), which may provide important information on clinical conditions or the efficacy of clinical treatments, other types of approaches are required.

Therefore, in this study, I investigated the temporal associations between depressive mood and behavioral dynamics in patients with MDD using ecological momentary assessment (EMA) [2]. I particularly examined the manner in which within-individual temporal changes in depressive mood scores co-varied with local statistics of locomotor activity around the recordings of self-reported symptoms. Furthermore, I compared such associations across patients with MDD and healthy subjects. The identification of differences and/or similarities in the psychobehavioral correlates between patients and healthy subjects is considered important because it may provide valuable insights into the pathogenic processes of MDD, thus leading to the early detection of the disease.

4.2. Methods

4.2.1. Subjects

The data were acquired from 14 patients with MDD [12 males (M)/2 females (F); age, 34.0 ± 5.7 years; age range, 22–42 years] and 43 healthy office workers (43M/0F; 40.7 ± 9.1 years; 23–58 years). The patients with MDD were outpatients of the Teikyo University Mizonokuchi Hospital, Kanagawa, Japan, and their locomotor activity data have been published elsewhere [126,127]. The patients who applied for participation in the study were interviewed and screened by a well-trained psychiatrist. The inclusion criteria were as follows: a diagnosis of MDD according to the Diagnostic and the Statistical Manual of Mental Disorders (DSM)-IV [124]; information on the current depressive episode; and age between 20 and 55 years. The exclusion criteria were: current substance abuse and other psychiatric diseases; lifetime history of schizophrenia or personality disorder; or severe physical illness. The 17-item Hamilton Depression Rating Scale (HDRS) [154] was also administered to all the patients [13.3 (mean) ± 2.9 (*SD*); range, 8.8–18.5]. The detailed profiles of the patients are summarized in Table 4-1.

The healthy subjects were full-time office workers at the University of Tokyo. They did not have a past or current diagnosis of any psychiatric diseases. Fifty-three subjects completed their recordings for the entire study period, but I excluded 10 subjects who rated a Beck Depression Inventory-II (BDI-II) score of 14 points or more (i.e., clinical cutoff level for depression) [155,156]. Thus, 43 subjects were analyzed in this study, and the mean score on BDI-II in this group was 5.9 ± 4.0 (range 0–13).

All subjects in this study were given a full explanation of the purposes and potential risks of the study by well-trained researchers. Subsequently, they signed an institutionally approved informed consent form. This study was approved by the research ethics committees of Teikyo University and

Table 4-1. Profiles of patients with major depressive disorder

| Patient No. | Gender [male(M)/female (F)] | Age | Employed [yes (Y)/no (N)] | Disease duration (month) | HDRS score (17-items) |
|-------------|--------------------------------|-----|------------------------------|-----------------------------|--------------------------|
| 1 | M | 42 | Y | 135 | 12 |
| 2 | M | 35 | N | 7 | 11.2 |
| 3 | F | 33 | N | 36 | 18.5 |
| 4 | M | 37 | N | 34 | 10 |
| 5 | M | 33 | N | 10 | 12.8 |
| 6 | M | 35 | N | 19 | 8.8 |
| 7 | M | 32 | N | 18 | 15.5 |
| 8 | M | 26 | Y | 69 | 13 |
| 9 | F | 22 | N | 22 | 12.4 |
| 10 | M | 33 | Y | 33 | 18.1 |
| 11 | M | 29 | Y | 7 | 10.8 |
| 12 | M | 38 | N | 21 | 13 |
| 13 | M | 40 | N | 2 | 16.8 |
| 14 | M | 41 | Y | 36 | 13 |

HDRS: Hamilton Depression Rating Scale.

the University of Tokyo.

4.2.2. Materials, procedures, and measurements

Recording of self-reported symptoms by ecological momentary assessment

An EMA technique [2] was used to record momentary symptoms in patients with MDD and healthy subjects. This approach allows us to address subjects' behaviors, psychological states, and physiological reactions at multiple time points as the individual experiences them in daily life. The collection of data in natural settings can enhance the validity of measurements, thus avoiding the pitfalls of retrospective recall, which highly distort self-reported data collection.

A small watch-type computer (Ruputer, ECOLOG, 42 g; Seiko Instruments Inc., Tokyo, Japan; Figure 2-1A) was used as an ED to record self-reported symptoms [36,39,128]. Patients with MDD

were requested to record their momentary symptoms by answering EMA questionnaires over the study period (37.43 ± 14.82 days; range, 18–67 days). EMA questionnaires prompted the patients to record their symptoms via a beep signal at randomly selected times within ± 36 min of the predefined times (6:00, 12:00, 18:00, and 24:00) during waking periods. Moreover, healthy subjects were instructed to complete the questionnaires at randomly selected times within ± 24 min of the predefined times (every four hours: 8:00, 12:00, 16:00, ...) during waking periods over seven consecutive days. In addition to these scheduled times, all subjects were also requested to register the time at which they woke up or went to bed as well as their momentary symptoms at that time (Figure 4-1).

The EMA questionnaires assessed subjective mood states and the intensity of physical symptoms (fatigue, sleepiness, pain, etc.) using a visual analog scale (0–100 with 5-point intervals). The mood states were rated using the Depression and Anxiety Mood Scale (DAMS) [120], which was developed to measure anxious and depressive moods as separately as possible. The DAMS comprises the following nine adjectives representing mood states: “vigorous”, “gloomy”, “concerned”, “happy”, “unpleasant”, “anxious”, “cheerful”, “depressed”, and “worried”. These nine items were used to calculate anxious and depressive mood scores (see 2.3.3 for details). In this study, I focused on the depressive mood because depressive symptoms are the most prominent feature of MDD, and mood changes are considered an important sign of the pathogenesis of the disease. In addition, consistent with my prior study (see the study in part 3) [76], I did not find significant associations of the other symptoms of depression, including anxious mood with local statistics of locomotor activity.

Assessment of locomotor activity

The watch-type device is also equipped with an activity monitor that has a performance that is analogous to that of the commercial actigraph (Ambulatory Monitors Inc., Ardsley, NY, USA), which

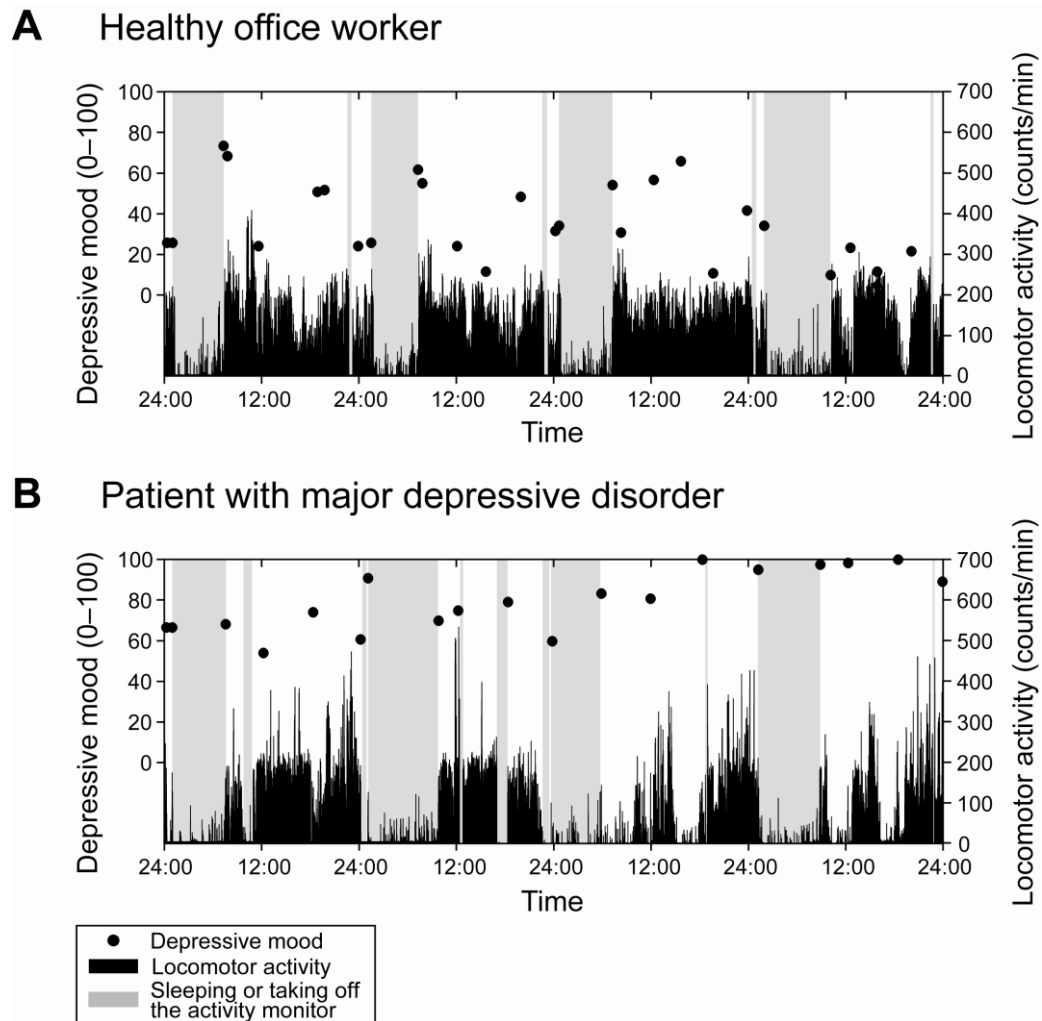


Figure 4-1. Fluctuations of depressive mood scores and locomotor activity. Filled circles indicate depressive mood scores (y-axis on the left side) recorded by ecological momentary assessment over four consecutive days in a healthy office worker (A) and a patient with major depressive disorder (B). Their simultaneously measured locomotor activities are also shown (y-axis on the right side). The periods during which they were sleeping or had taken off the device are shaded in gray.

is used widely in the clinical field [21,126,127,128]. The sensor for assessing locomotor activity is a uni-axial piezo-electronic accelerometer that is capable of detecting small changes in bodily acceleration (≥ 0.01 G/rad/s), which enables the registration of even slight movements in daily life. All subjects were instructed to wear this device on the wrist of their respective non-dominant hand throughout the study period, except while bathing, showering, performing rigorous exercises, or any other activity likely to damage the device. In this study, zero-crossing counts accumulated for every 1

min were used as locomotor activity data (Figure 4-1). Locomotor activity data collected during periods in which the subjects were not wearing the device or sleeping were excluded from the analysis.

4.2.3. Data analysis

Local statistics of locomotor activity

I focused on the first- and third-order statistical moments (i.e., mean and skewness) of locomotor activity because the combination of these statistics can well characterize the intermittent or bursty nature (i.e., reduced activity levels and occasional bursts leading to a positively skewed probability distribution) of the data [76]. These local statistics were calculated from locomotor activity data with a length of 60 min centered around each EMA recording. The effects of the data length (size) and the location against the timing of EMA recording were examined later.

4.2.4. Statistical analysis

I adopted a multilevel modeling approach [105,157,158] using SAS Proc Mixed (SAS 9.2, SAS Institute Inc., Cary, NC) because the present study produced a hierarchically structured data set in which the EMA for depressive mood and the corresponding local statistics of locomotor activity were nested within subjects [107]. A $p < 0.05$ was considered significant.

Descriptive statistics of recording profiles

The mean level of momentary symptoms was estimated using the so-called null model, in which each momentary symptom is the dependent variable with no predictor. To test the group differences in momentary symptoms, I adopted a 2-level mixed MANOVA model using each momentary symptom

as the dependent variable and the type of group (i.e., MDD or healthy individuals) as the predictor. Moreover, I performed the same analysis for the local statistics of locomotor activity.

Statistical model for depressive mood

First, I identified the statistical model that described the associations between depressive mood scores and the local statistics of locomotor activity using a combined data set from the healthy and MDD groups. I compared all possible multilevel models that consisted of a linear combination of a subset of both local statistics and their interaction. In addition, I examined both fixed and random effects for each predictor in each model. All variance components were assumed to follow a normal distribution with zero mean. I used the deviance statistic to compare the goodness of fit of the models with nested or inclusive relations, in which all predictors in one model were a subset of the predictors in another model [105].

After identifying the statistical model that best fitted the data described above, I also examined the robustness of this model against the choice of the data length and temporal location that were used to derive the local statistics of locomotor activity. Particularly, I varied the data length from 10 to 120 min with an increment of 10 min centered around each EMA recording. I also varied the center location of the data from -30 to 30 min of the EMA recording, with the data length set at 60 min.

Group differences in associations between depressive mood and locomotor activity

To examine the differences in the associations between depressive mood scores and the local statistics of locomotor activity across the MDD and healthy groups, I added a categorical variable representing the type of group into the parameters of the model identified (i.e., the intercept and slopes of predictors). This test allowed us to examine whether the parameter values differed between groups

and provided the estimates of the model parameters for each group. These estimates were used in the cross-validation study.

Cross-validation of the statistical model across groups

As shown below, I did not find any group differences in the model parameters, with the exception of the intercept. This indicates that the only difference in the associations between groups was an overall level of depressive mood, which suggests that it is possible to estimate changes in mood scores in one group using the model derived from the other group and vice versa, once the adjustment for the intercept is properly conducted. The cross-validity of the derived models was particularly examined as follows. I estimated the depressive mood scores of healthy subjects using the model with the parameter values for the MDD group provided by the analysis described above. To estimate the scores, I substituted the local statistics of locomotor activity of a healthy subject into the model of the MDD group, in which the local statistics were evaluated from the data with a length of 60 min centered around EMA recordings. All random terms in the model were set to zero because of their definition. In addition, I estimated the depressive mood scores of patients with MDD using the model of the healthy group. The cross-validity was evaluated by examining the correlation coefficient between estimated and self-reported depressive mood scores via a 2-level mixed MANOVA model using depressive mood scores as the dependent variable and the type of score (i.e., estimated or self-reported score) as the predictor [39].

4.3. Results

4.3.1. Recording profiles

It was obtained 1921 [137.2 (mean) \pm 63.6 (*SD*) per person] EMA recordings from patients with MDD and 1781 recordings (41.4 ± 5.0) from healthy subjects. If simultaneous locomotor activity data were not acquired properly because of trouble with or removal of the device, I excluded the data from the analysis. Finally, I obtained 767 (54.8 ± 23.3) sets of simultaneous recordings of EMA and locomotor activity data from patients with MDD and 946 (22.0 ± 4.7) such recordings from healthy subjects.

Table 4-2 summarizes the mean levels of subjective symptoms and the local statistics of locomotor activity used for the analysis. The mean levels of all symptoms in patients with MDD were significantly higher than those observed in healthy subjects ($p < 0.001$). The mean activity levels of patients with MDD were significantly lower than those of healthy subjects ($p < 0.001$), whereas the values of skewness were significantly higher in the patients than in controls ($p < 0.001$). These

Table 4-2. Mean levels of momentary symptoms and local statistics of locomotor activity

| | mean(<i>SE</i>) | | <i>p</i> value |
|---|-------------------|--------------|----------------|
| | MDD | Healthy | |
| Momentary symptoms | | | |
| Depressive mood | 58.95(2.73) | 41.64(1.59) | <0.001 |
| Anxious mood | 58.68(4.15) | 39.03(2.41) | <0.001 |
| Fatigue | 56.59(5.04) | 31.12(2.90) | <0.001 |
| Local statistics of locomotor activity | | | |
| Mean | 111.59(5.08) | 132.61(3.08) | <0.001 |
| Skewness | 0.47(0.07) | 0.11(0.04) | <0.001 |

SE: standard error. MDD: major depressive disorder. All local statistics of locomotor activity were evaluated from the data collected within 60 min centered around the ecological momentary assessment recording.

differences in the local statistics of locomotor activity indicate the increased intermittency of behavior dynamics, which is phenomenologically compatible with the recent findings of alterations of scale-invariant statistics in patients with MDD derived from long-term (>7 days) locomotor activity data [126]. This indicates that the increased intermittency observed in patients can also be captured by alterations in the local statistics, particularly the combination of decreased mean activity levels and higher positive values of skewness.

4.3.2. Determination of statistical model for depressive mood

The statistical model identified from the combined data set from both groups was a linear combination of the local mean, skewness, and their interaction as predictors, in which the local mean and intercept had a random effect: Depressive mood scores $_{ij} = \gamma_{00} + \gamma_{10} (\text{Mean}_{ij}) + \gamma_{20} (\text{Skewness}_{ij}) + \gamma_{30} (\text{Mean}_{ij} \times \text{Skewness}_{ij}) + \zeta_{0i} + \zeta_{1i} (\text{Mean}_{ij}) + \varepsilon_{ij}$, where Depressive mood scores $_{ij}$ indicates a score of depressive mood at the j th recording for the i th subject; Mean_{ij} and Skewness_{ij} are the local mean and skewness of locomotor activity corresponding to the j th EMA recording for the i th subject, respectively; γ_{00} is the average intercept across all subjects; γ_{10} , γ_{20} , and γ_{30} are the average regression slopes across all subjects for each predictor; the random terms ζ_{0i} and ζ_{1i} are the between-individual residuals; and ε_{ij} is the within-individual residual. The coefficients of all the predictors were significant (Table 4-3), with a positive value ($\gamma_{20} = 1.82 \pm 0.70$, $p = 0.009$) for skewness and negative values for both the mean ($\gamma_{10} = -0.03 \pm 0.01$, $p = 0.008$) and interaction ($\gamma_{30} = -0.01 \pm 0.01$, $p = 0.030$). These results indicate the presence of a significant association between increased intermittency of locomotor activity and worsening of self-reported depressive mood. The negative coefficient of interaction term possibly reflects situations in which the depressive mood becomes worse with both lower activity levels and positively larger skewness; thus, more frequent episodes of bursts are observed.

Table 4-3. Statistical model for depressive mood and group differences regarding model parameters

| | Coefficient(<i>SE</i>) | <i>F</i> value | <i>p</i> value |
|---|--------------------------|---------------------|----------------|
| Statistical model | | | |
| Intercept (γ_{00}) | 50.11(2.45) | | <0.001 |
| Coefficient of the local mean (γ_{10}) | -0.03(0.01) | $F_{1,1653} = 7.01$ | 0.008 |
| Coefficient of the local skewness (γ_{20}) | 1.82(0.70) | $F_{1,1653} = 6.81$ | 0.009 |
| Coefficient of the local mean \times local skewness (γ_{30}) | -0.01(0.01) | $F_{1,1653} = 4.70$ | 0.030 |
| Group difference | | | |
| Intercept (γ_{00}) | | $F_{1,55} = 18.18$ | <0.001 |
| MDD | 64.41(3.70) | | <0.001 |
| Healthy | 44.41(2.71) | | <0.001 |
| Coefficient of the local mean (γ_{10}) | | $F_{1,1650} = 1.02$ | 0.312 |
| MDD | -0.05(0.02) | | 0.013 |
| Healthy | -0.03(0.02) | | 0.133 |
| Coefficient of the local skewness (γ_{30}) | | $F_{1,1650} = 0.23$ | 0.631 |
| MDD | 1.79(0.79) | | 0.023 |
| Healthy | 0.91(1.67) | | 0.587 |
| Coefficient of the local mean \times local skewness (γ_{30}) | | $F_{1,1650} = 0.18$ | 0.671 |
| MDD | -0.01(0.01) | | 0.079 |
| Healthy | -0.01(0.01) | | 0.585 |

SE: standard error. MDD: major depressive disorder. The statistical model is as follows: Depressive mood scores_{*ij*} = $\gamma_{00} + \gamma_{10}$ (Mean_{*ij*}) + γ_{20} (Skewness_{*ij*}) + γ_{30} (Mean_{*ij*} \times Skewness_{*ij*}) + $\zeta_{0i} + \zeta_{1i}$ (Mean_{*ij*}) + ϵ_{ij} .

I demonstrated the robustness of the model identified against the choices of the data length and temporal location that were used to calculate the local statistics of locomotor activity. Figure 4-2 shows the dependency of the coefficients of three predictors on the different values of choices. All coefficients were consistently significant over a broad range of location (Figure 4-2A, C, and E) and data length (Figure 4-2B, D, and F) values, which was indicative of the robustness of the model identified.

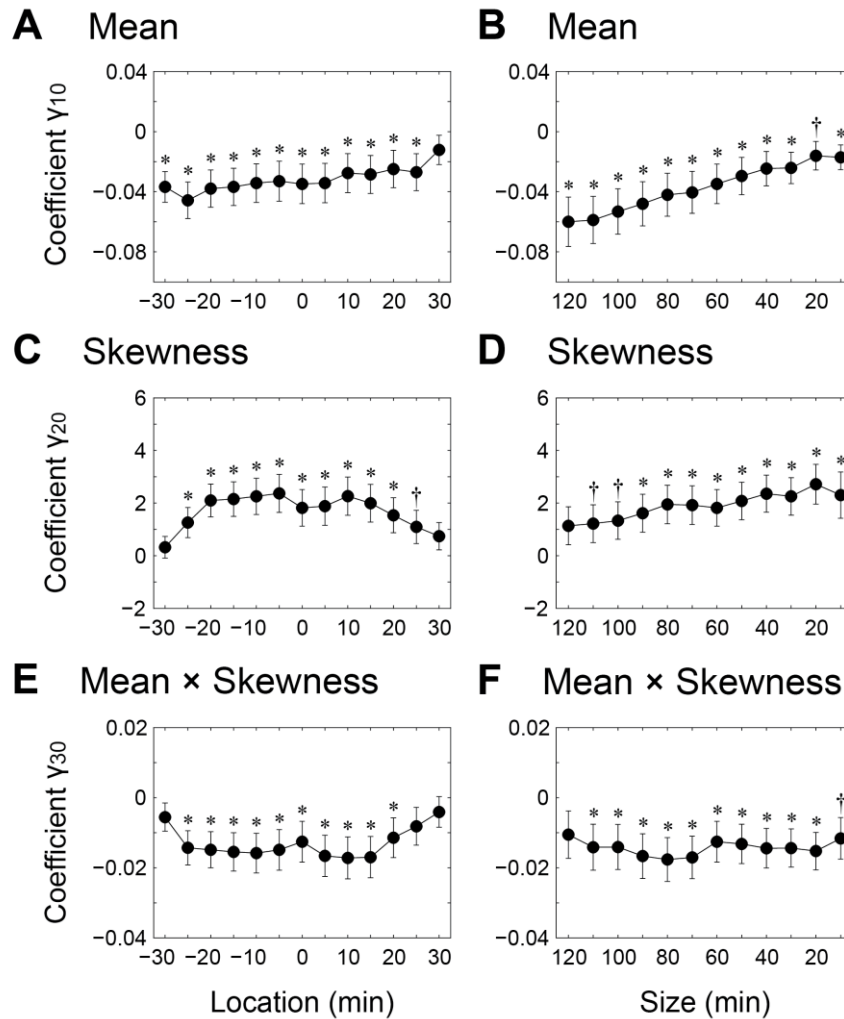


Figure 4-2. Dependency of model coefficients on different locations and data lengths. Coefficient of the predictors: (A) Mean (γ_{10}), (C) Skewness (γ_{20}), and (E) Mean \times Skewness (γ_{30}) of the model identified as a function of the location. In these panels, the data length was fixed at 60 min. The same is shown in panels (B), (D), and (F), with the exception of the dependency on the data length. Here, the center of the data used for the derivation of the statistics was fixed at the timing of the ecological momentary assessment recording. Error bars indicate the standard error of the mean. The asterisks and daggers indicate that the model coefficients are significantly different from zero at $p < 0.05$ and 0.10 , respectively.

4.3.3. Group differences in associations between depressive mood and locomotor activity

There were no significant differences in the coefficient of all predictors, with the exception of the model intercept (Table 4-3). This indicates that identical relationships between changes in mood and behavior exist across patients with MDD and healthy subjects, although the patients had higher levels

of overall depressive mood scores, as represented by the higher intercept values.

Table 4-3 summarizes the estimates of model coefficients for both groups. All the coefficients in patients with MDD were significant, with the exception of the interaction term with a significant positive value ($\gamma_{20} = 1.79 \pm 0.79$, $p = 0.023$) for skewness and negative values for both the mean ($\gamma_{10} = -0.05 \pm 0.02$, $p = 0.013$) and their interaction ($\gamma_{30} = -0.01 \pm 0.01$, $p = 0.079$). In contrast, none of the coefficients estimated in healthy subjects were significant ($\gamma_{20} = 0.91 \pm 1.67$, $p = 0.587$ for skewness; $\gamma_{10} = -0.03 \pm 0.02$, $p = 0.133$ for the mean; and $\gamma_{30} = -0.01 \pm 0.01$, $p = 0.585$ for the interaction), whereas the signs of all coefficients were consistent with those estimated for the MDD group. The difference in the significance of the model parameters between groups may have been caused by the difference in within-individual sample sizes (on average: 54.8/person for the MDD group and 22.0/person for the healthy group). However, even these parameters for the healthy group fit well in the cross-validation study, as shown below.

4.3.4. Cross-validation of statistical model across groups

The correlation between self-reported depressive mood in patients and estimated scores based on the model for the healthy group was significant (correlation coefficient $r = 0.21$; $p < 0.001$). Regarding the depressive mood scores of healthy subjects, the correlation coefficient was also significant ($r = 0.10$, $p < 0.001$). These results indicate the cross-validity of the models identified or the presence of shared psychobehavioral correlates across groups.

4.4. Discussion

In this study, I examined psychobehavioral correlates between momentary depressive mood and behavioral dynamics in patients with MDD. I particularly evaluated the associations between changes in momentary self-reported depressive mood scores and the local statistics of locomotor activity simultaneously measured. The statistical model that described the associations showed increased intermittency of behavioral dynamics with worsening of depressive mood. This result suggests that it is possible to estimate momentary depressive mood from the objective measures of locomotor activity, thus leading to continuous monitoring of the pathogenic processes and pathological changes of MDD. Furthermore, I confirmed the cross-validity of the model describing within-individual associations across healthy subjects and patients with MDD. The existence of shared psychobehavioral correlates between groups implies that changes in momentary depressive mood in healthy individuals, like those recorded for clinically depressed patients, can also be estimated from locomotor activity data. However, I also found the group difference in the overall level of depressive mood score (i.e., the intercept), suggesting that the model for healthy individuals cannot be simply used to predict depressive mood levels of the patients. Such discontinuity in psychobehavioral relations between healthy subjects and the patients might have implications for the pathogenic processes, hence the prevention and/or early detection of MDD. Continuous monitoring of psychobehavioral dynamics for the borderline subjects would be expected to provide more detailed pathogenic processes in the future research.

The intermittent and bursty nature of human behavioral dynamics is now receiving attention in various scientific fields [159,160,161,162,163]. Recently, increased intermittency of locomotor activity was found in patients suffering from MDD [126] and schizophrenia [129], which was indicative of more frequent episodes of slowing down or cessation of movement while having

occasional bursts in these patients compared with healthy subjects (e.g., days 3–4 of Figure 4-1B compared with Figure 4-1A). Furthermore, it was proposed a possible theoretical model for intermittent behavioral dynamics and its alterations based on a priority stochastic queuing theory [164]. According to this model, the increased intermittency observed in patients with MDD can be explained by a strategic change in decision making to initiate actions with preferential selectivity to demands (cues) with higher priority. Patients with MDD tend to react only to higher demands and/or stimuli (generating occasional activity bursts) and stay quiet for most of the time (resulting in reduced activity). The results of this study further suggest that such a strategic change can be associated with alterations in momentary depressive mood, leading to increased local intermittency in locomotor dynamics.

The advancement of information and communication technologies (ICTs) has led to the recent rigorous development of healthcare monitoring systems that often combine mobile technologies [123,165,166] and are expected to play a considerable role in early detection, management, and treatment of psychiatric disorders, including MDD, bipolar disorder, and schizophrenia [165,166]. To successfully establish such systems, many essential elements have been under study. For example, the concept of ecological momentary interventions (EMIs), in which real-time interventions are delivered to individuals during their everyday lives in natural settings, is a core elemental technology that is used for novel treatments of these diseases [167]. In contrast, the sign-contingent EMA, which is an assessment that is triggered by a warning “sign” related to the disorders, is considered useful for detecting early signs of psychiatric disorders and their pathological transitions [7,168]. However, the development of objective, reliable, and, more importantly, real-time biobehavioral markers for psychiatric disorders is necessary to establish these important elements; thus, the demand for such markers has grown recently. In this context, the current study has a great potential to contribute to the

establishment of healthcare systems by providing an objective and real-time biobehavioral marker, i.e., a statistical model for depressive mood based on behavioral dynamics. In addition, these findings serve as useful models for continuous monitoring and estimation of depressive mood as well as fundamental techniques for the realization of EMIs and sign-contingent EMA.

One potential limitation of this study was that the correlation coefficients obtained in the cross-validation of the identified model were small, although the correlations themselves were significant. These low correlation values may be because of wide individual differences, such as lifestyle, pathological conditions, and effects of antidepressant medication. However, these individual effects may be minimized by optimization methods such as support vector machine [169] or neural network approaches [170]. For example, the personalization of the model structure and its parameters using the data of the first few weeks could minimize the individual effects and improve the ability to estimate depressive mood scores. In fact, I could improve the estimation of the depressive mood scores merely by personalizing the model parameters using simple linear models employed in the present study. Figure 4-3 shows typical cases in which the estimation was highly improved by this type of personalization. The parameters of the model derived in this study were optimized individually using data collected at one week in the early part of the measurement. Subsequently, the depressive scores in another week in the later part of the study were estimated using personalized parameters. In these patients, the correlation coefficients between self-reported and estimated depressive mood scores were considerably higher [$r = 0.80$ ($p = 0.002$) for Figure 4-3A and $r = 0.74$ ($p = 0.004$) for Figure 4-3B] than those calculated for the overall cross-validation. Although I was unable to perform this optimization procedure for the data from all patients because this requires relatively long-term measurements (>2 weeks), I confirmed considerable improvements in the estimation in six patients who had this condition with correlation coefficients ranging from 0.48 to 0.80. Other

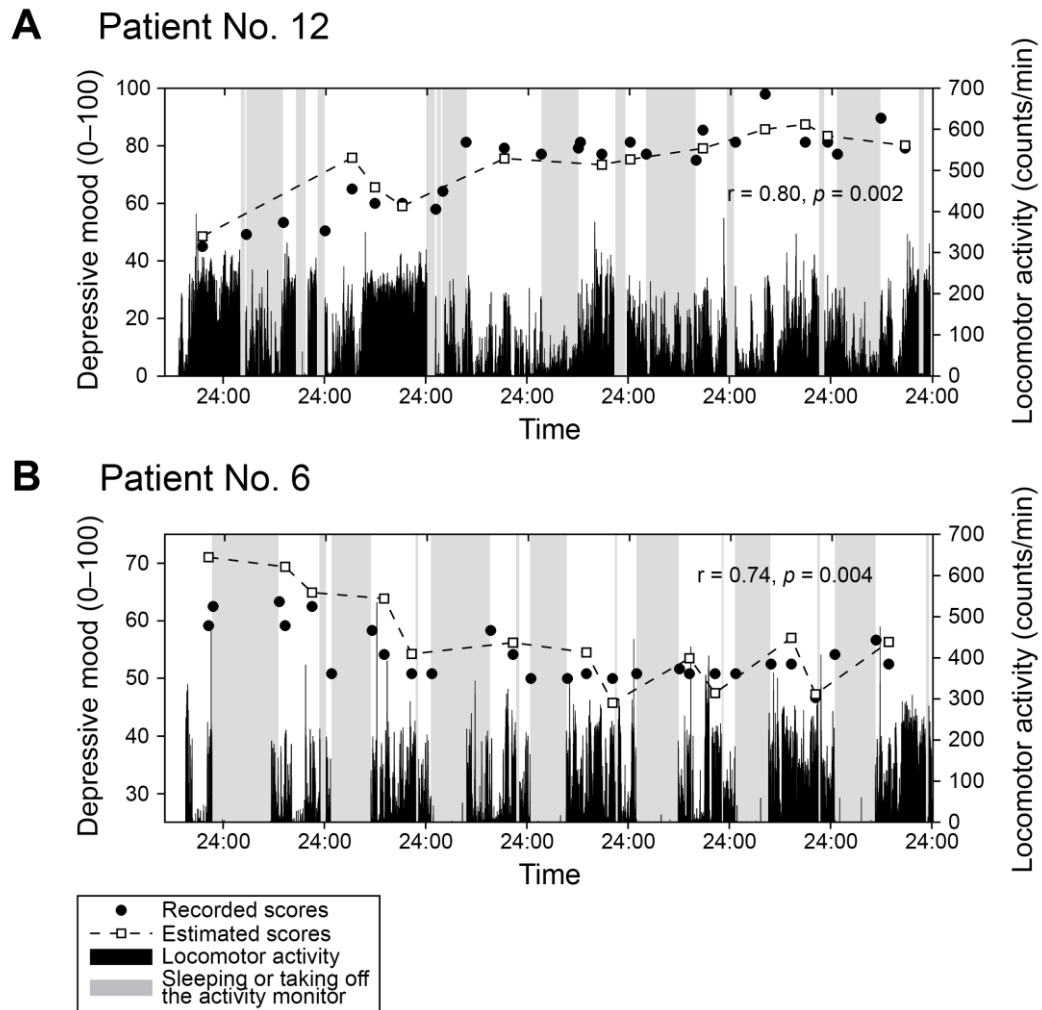


Figure 4-3. Estimation of depressive mood scores using personalized models. Open squares indicate the estimates of depressive mood scores (y-axis on the left side) by substituting the local statistics of locomotor activity (within 60 min centered around the ecological momentary assessment recording) into the statistical model with personalized parameter values. The personalized parameter values for subject No. 12 (A) were: $\gamma_{00} = 83.78$, $\gamma_{10} = -0.08$, $\gamma_{20} = 1.17$, and $\gamma_{30} = -0.01$, whereas those for subject No. 6 (B) were: $\gamma_{00} = 70.38$, $\gamma_{10} = -0.15$, $\gamma_{20} = 4.08$, and $\gamma_{30} = -0.06$. The correlation coefficients r and their p values between the estimated and self-reported scores were 0.80 and 0.74, respectively. Note that the estimation of depressive mood was only made for periods in which the subjects were awake and wearing the device.

sophisticated optimization methods would probably highly improve the estimation.

The other weaknesses of this study included the effects of medications on behavioral dynamics and the small sample size. Thus, the generalization of the findings of this study will require a large population study using drug-free patients. Another issue that should be addressed is the ceiling effect.

Although patients with MDD showed a level of variability in depressive mood scores that was comparable with that of healthy subjects, four patients occasionally reported maximum scores during the study period. Because of the low rate (2.76 %) of such data points, I believe that the ceiling effect was limited in this study; however, there is a possibility that the associations observed were slightly distorted by such an effect.

In conclusion, I demonstrated the presence of associations between momentary depressive mood and behavior dynamics and showed their cross-validity across patients with MDD and healthy subjects. These results suggest that it is possible to objectively estimate a momentary depressive mood based on changes in physical activity, thus leading to continuous monitoring of the pathogenic processes and pathological states of MDD.

PART 5: GENERAL DISCUSSION

5.1. Summary of findings

The purpose of this thesis was to examine the utility of ecological momentary assessment (EMA) as a method for capturing momentary symptoms. I introduced the conceptual, historical, applicative, methodological, and statistical aspects of EMA in part 1. Based on the comprehensive discussion of the rationale behind EMA, I address what I believe important and timely research questions with regard to the utility of EMA in parts 2–4. The findings set out in those three parts are summarized below.

In part 2, a full systematic comparison of EMA and the day reconstruction method (DRM) was conducted for fatigue and mood states in healthy adults. The findings showed no significant differences between the mean or the variability of EMA and DRM estimated over the monitoring period. However, correlations between EMA and DRM were low, albeit statistically significant. This indicates that although the overall mean and variability of EMA recordings may be accessible with DRM, detailed changes over time in momentary fatigue and mood states cannot be retrieved using DRM. In addition, I found that reconstructed depression through DRM could be biased when subjects remembered whether their behavior was active or inactive. Although prior studies have assured that recalling the context of daily activities with DRM is helpful in reconstructing subjective symptoms without recall bias [9,11,122], there may be systematic biases in DRM assessment, and its biases may be related to certain situational information. These findings indicate that EMA may be the best way to measure a person's diurnal time course (i.e., detailed changes over time) with regard to fatigue and mood states.

In part 3, I first examined the variations in momentary symptoms by validating the associations

among self-reported symptoms measured simultaneously (depressive mood, anxious mood, and fatigue), and then investigated covariant properties between the symptoms (especially, depressive mood) and local statistics of locomotor activity as the external objective criteria obtained continuously. The results showed convergent associations by showing positive correlations among momentary symptoms. Also, the increased intermittency of locomotor activity, characterized by a combination of reduced activity with occasional bursts, appeared concurrently with the worsening of depressive mood. Further, this association remained statistically unchanged across groups regardless of group differences in age, lifestyle, and occupation. These results indicate that temporal variations in momentary symptoms are not random but reflect the underlying changes in psychophysiological variables in daily life. In other words, momentary symptoms recorded using EMA are valid data reflecting physiological states.

In part 4, I investigated psychobehavioral correlates, particularly the statistical associations between momentary depressive mood and behavioral dynamics measured objectively, in patients with major depressive disorder (MDD) and in healthy subjects. The statistical model established indicated that worsening of depressive mood was associated with increased intermittency of locomotor activity, as characterized by a lower mean and higher skewness. The model was cross-validated across groups, suggesting that the same psychobehavioral correlates are shared by both healthy subjects and patients with MDD, although the latter had significantly higher mean levels of depressive mood scores. These findings suggest the presence of robust as well as common associations between momentary depressive mood and behavioral dynamics in healthy individuals and patients with depression, which indicates the value of the continuous monitoring of the pathogenic processes (from healthy states) and pathological states of MDD by using EMA.

In summary, the utility of EMA as a method to capture momentary symptoms was confirmed in

both healthy and diseased samples.

5.2. Implication of findings

The findings of this thesis show that EMA is useful for assessing self-reported symptoms. Based on these findings, it can be concluded that the EMA approach can be used to improve our understanding of the basic nature of fluctuations in patterns of self-reported symptoms. In addition, an EMA approach incorporating information and communication technologies (ICTs) will enable a more detailed description of these patterns and how they related to other variables such as physiological states and thus may contribute to a more refined understanding of psychiatric disorders including their pathogenic processes, prevention, and treatment.

5.2.1. Utility of EMA in field of psychiatry

The first step toward prevention and treatment of psychiatric disorders is to thoroughly understand the symptoms of the disorders [18]. Frequent/repeated assessments of momentary symptoms make it possible to capture within-individual variability and changes in the symptoms of psychiatric disorders, e.g., MDD, as shown in part 4. Prior to EMA procedures, subjects were usually asked to reconstruct symptoms, e.g., they were asked how their depressive mood fluctuated over a day, and thus actual assessment of variability was uncommon. EMA allows us to explore the development trajectories of the disorders and identify factors that are predictive of these trajectories. Also, as discussed in part 2, such fluctuations in the trajectories cannot be captured by using DRM.

In addition, patterns of self-reported symptoms often arise relating to time of day or other factors, and these patterns give clues both to the function of the disorders in patients' lives and potential strategies to alleviate the symptoms [18]. Many studies [28,29,30,31,34,88,91,111] using EMA, including the studies in parts 3 and 4, indicate that the diurnal patterns of self-reported symptoms are associated with physiological states. The patterns are also correlated with social and environmental

situations [38,52]. In contrast to standard clinical interviews or single measurements, EMA offers a tool which can be used to examine antecedents or possible triggers of psychiatric disorders. In other words, EMA can be effectively used to assess related symptoms with other related factors immediately before and just after disorders (e.g., panic attacks [72], binge eating [65]), which give us important insights into pathogenic processes and prevention of psychiatric disorders.

When we think of patients with psychiatric disorders, we probably imagine that their related symptoms are constantly and chronically high. However, momentary symptoms actually fluctuate over the course of a day or week despite their high average level (see Figure 4-1B and [171]). EMA is well suited to capturing this substantial moment-to-moment variation around the average levels, and makes precise evaluation of treatment possible. For example, it can be used to evaluate the effects of medication over time and also to identify clues (e.g., physiological states or social and environmental situations) for diurnal variation in momentary symptoms related to psychiatric disorders.

5.2.2. Using EMA with information and communication technologies

Recent advances in ICTs have given clinicians and researchers powerful new tools that can be used in health-related data collection. A person's momentary symptoms fluctuate according to his or her internal state (e.g., physiological state) or the external situation (e.g., behavioral episode), and vice versa. However, the relationships and causality among these variables remain unknown. Up until about 30 years ago, there were not even any tools to investigate this issue. In the last few decades, however, outstanding progress has been made in capturing psychological, physiological, and behavioral variables in (close to) real time with sophisticated monitoring devices. Today, many researchers on the behavioral sciences and other scientific disciplines use EMA with ICTs for their work, which has led to improved quality and quantity of data, as follows [7].

Most sampling for EMA data collection uses electronic devices with small screens, and answers are stored by pushing a button or touching the screen with time information. This has improved the quality of data collection in terms dealing with the problem of faked compliance and via automated storage of the data set (i.e., the devices eliminate errors that can arise when transferring data from paper to a computer). In the coming years, the use of electronic devices (e.g., activity monitors as used in part 3 and 4) to comfortably, continuously, and automatically collect data will accelerate. Such devices, worn on the body or placed in the environment, will record information about a subject's behavior, location, and physiological states without burdening subjects with the need to self-report. In addition, emerging electronic devices will make "context-sensitive prompting" possible, where questions are automatically triggered based on the subject's behavior, location, physiological states, past responses, and social interactions. Furthermore, emerging electronic devices will allow us to collect and store an unprecedented amount of longitudinal data. Algorithms that analyze this data will be able to estimate and predict target variables based on a subject's prior behavior or self-reported symptoms, cognitions, or social interactions (see the discussion section in part 4 and Figure 4-3). All these functions will make EMA data richer and more reliable, and thus enable researchers to make further advances in many scientific fields.

5.3. Perspectives for future study

A number of questions remain to be resolved, and I would like to finish this thesis by outlining future directions for my research. Although I found that detailed changes over time in momentary symptoms are not retrieved by recalled assessment (i.e., DRM) in part 2, the meaning of this finding should be discussed further. Researchers who support a real-time point of view believe that momentary assessment is a better way of capturing what subjects have experienced and that recalled assessment is biased by memory processes. However, the well-known colonoscopy study [172] demonstrated that whether or not patients will present for colonoscopies again in the future can be predicted by their recalled pain immediately after the original colonoscopy. In other word, the impression left in the memory predicted future behavior in this case. The logic behind this issue should be carefully considered according to the experience we are trying to capture in future studies.

In addition, I could not extend the basis of the study in parts 3 and 4 to include exploration into the mechanisms that caused the associations between momentary symptoms and locomotor dynamics, but I believe that future work should address these mechanisms in more detail. As I described above, with advances in ICTs, there has been a growing international trend for the development of intelligent monitoring systems [123,165,166] aimed at the prevention and treatment of psychiatric disorders. The success of such systems, however, is considered to depend crucially upon the presence of objective, reliable, and real-time biomarkers for psychiatric disorders with clearly revealed mechanisms. While my thesis successfully provided a bio-behavioral marker based on locomotor activity data, other types of biomarker are necessary to examine the mechanisms multi-dimensionally by providing richer information on physiological conditions, e.g., cardiovascular functions as a potential biomarker. Interpersonal and environmental experiences should also be considered in order to reveal the mechanisms.

Finally, ethical and privacy concerns related to EMA data collection should be considered in future studies [7,19]. A current electronic device implementing EMA is capable of recording a person's location, activities, social interactions, and physiological/psychological states, and future devices will be able to generate even more data about people than the devices used in current studies. How such data (e.g., data on a person's health status) is managed can be a concern. This concern could be addressed by using password-protected devices and protocols, using a stand-alone computer for data management, and ethical training for all people who have access to the data. Other issues are more complex, such as how to deal with data revealing subjects' health problems (e.g., heart rate), crises or emergencies (e.g., suicide), and illegal activities (e.g., substance use). In addition, some devices using EMA may unintentionally capture data about non-consenting people who interact with subjects. Furthermore, researchers in EMA studies monitor not only logs of emails or texts but also their content. This last concern could be addressed by analyzing data immediately on the device, without saving the data itself, but further studies are needed to identify ways of addressing the other ethical and privacy concerns.

5.4. Conclusion

Psychologists and survey methodologists have made considerable progress in understanding the underlying mechanisms of self-reports for subjective symptoms. Despite the progress that has been made, however, how to measure the actual states of subjective symptoms is not yet fully understood. This thesis revealed the utility of EMA as a method for capturing subjective symptoms by investigating the association of momentary symptoms with reconstructed symptoms or physiological states. The present thesis contributes to the objective and rigorous assessment of subjective symptoms using EMA, which should help promote mental health and reduce the risk of psychiatric disorders.

REFERENCES

1. Robinson MD, Clore GL (2002) Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychol Bull* 128: 934-960.
2. Stone AA, Shiffman S (1994) Ecological momentary assessment (EMA) in behavioral medicine. *Ann Behav Med* 16: 199-202.
3. Shiffman S, Stone AA, Hufford MR (2008) Ecological momentary assessment. *Annu Rev Clin Psychol* 4: 1-32.
4. Bradburn NM, Rips LJ, Shevell SK (1987) Answering autobiographical questions: the impact of memory and inference on surveys. *Science* 236: 157-161.
5. Robinson MD, Clore GL (2002) Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *J Pers Soc Psychol* 83: 198-215.
6. Pickering TG, James GD, Boddie C, Harshfield GA, Blank S, et al. (1988) How common is white coat hypertension? *JAMA* 259: 225-228.
7. Stone AA, Shiffman S, Atienza AA, Nebeling L (2007) *The science of real-time data capture: self-reports in health research*. Oxford; New York: Oxford University Press.
8. Robinson JP, Godbey G (1997) *Time for life: the surprising ways Americans use their time*. University Park, Pa.: Pennsylvania State University Press.
9. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA (2004) A survey method for characterizing daily life experience: the day reconstruction method. *Science* 306: 1776-1780.
10. Martyn KK, Belli RF (2002) Retrospective data collection using event history calendars. *Nurs Res* 51: 270-274.
11. Stone AA, Schwartz JE, Schkade D, Schwarz N, Krueger A, et al. (2006) A population approach to the study of emotion: diurnal rhythms of a working day examined with the Day Reconstruction Method. *Emotion* 6: 139-149.
12. Barker RG, Associates (1978) *Habitats, environments, and human behavior*. San Francisco: Jossey-Bass.
13. Korotitsch WJ, Nelson-Gray RO (1999) An overview of self-monitoring research in assessment and treatment. *Psychol Assess* 11: 415-425.
14. Schlundt DG, Johnson WG, Jarrell MP (1985) A naturalistic functional-analysis of eating behavior in bulimia and obesity. *Adv Behav Res Ther* 7: 149-162.
15. Romanczyk RG, Tracey DA, Wilson GT, Thorpe GL (1973) Behavioral techniques in the treatment of obesity: a comparative analysis. *Behav Res Ther* 11: 629-640.
16. Sieck WA, McFall RM (1976) Some determinants of self-monitoring effects. *J Consult Clin Psychol* 44: 958-965.
17. Csikszentmihalyi M, Larson R (1987) Validity and reliability of the Experience-Sampling Method. *J Nerv Ment Dis* 175: 526-536.
18. Hektner JM, Schmidt JA, Csikszentmihalyi M (2007) *Experience sampling method: measuring the quality of everyday life*. Thousand Oaks, Calif.; London: Sage.
19. Trull TJ, Ebner-Priemer U (2013) Ambulatory assessment. *Annu Rev Clin Psychol* 9: 151-176.
20. Turner JR, Ward MM, Gellman MD, Johnston DW (1994) The relationship between laboratory and ambulatory cardiovascular activity: current evidence and future directions. *Ann Behav Med* 16: 12-23.
21. Teicher MH (1995) Actigraphy and motion analysis: new tools for psychiatry. *Harv Rev*

- Psychiatry* 3: 18-35.
22. Teicher MH, Glod CA, Magnus E, Harper D, Benson G, et al. (1997) Circadian rest-activity disturbances in seasonal affective disorder. *Arch Gen Psychiatry* 54: 124-130.
 23. Burton C, McKinstry B, Szentagotai Tatar A, Serrano-Blanco A, Pagliari C, et al. (2013) Activity monitoring in patients with depression: a systematic review. *J Affect Disord* 145: 21-28.
 24. Indic P, Murray G, Maggini C, Amore M, Meschi T, et al. (2012) Multi-scale motility amplitude associated with suicidal thoughts in major depression. *PLoS ONE* 7: e38761.
 25. Walther S, Hugli S, Hofle O, Federspiel A, Horn H, et al. (2012) Frontal white matter integrity is related to psychomotor retardation in major depression. *Neurobiol Dis* 47: 13-19.
 26. Volkers AC, Tulen JH, van den Broek WW, Bruijn JA, Passchier J, et al. (2003) Motor activity and autonomic cardiac functioning in major depressive disorder. *J Affect Disord* 76: 23-30.
 27. Berle JO, Hauge ER, Oedegaard KJ, Holsten F, Fasmer OB (2010) Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. *BMC Res Notes* 3: 149.
 28. Robles TF, Shetty V, Zigler CM, Glover DA, Elashoff D, et al. (2011) The feasibility of ambulatory biosensor measurement of salivary alpha amylase: relationships with self-reported and naturalistic psychological stress. *Biol Psychol* 86: 50-56.
 29. van Eck M, Berkhof H, Nicolson N, Sulon J (1996) The effects of perceived stress, traits, mood states, and stressful daily events on salivary cortisol. *Psychosom Med* 58: 447-458.
 30. Smyth J, Ockenfels MC, Porter L, Kirschbaum C, Hellhammer DH, et al. (1998) Stressors and mood measured on a momentary basis are associated with salivary cortisol secretion. *Psychoneuroendocrinology* 23: 353-370.
 31. Steptoe A, Gibson EL, Hamer M, Wardle J (2007) Neuroendocrine and cardiovascular correlates of positive affect measured by ecological momentary assessment and by questionnaire. *Psychoneuroendocrinology* 32: 56-64.
 32. Boland E, Monsod T, Delucia M, Brandt CA, Fernando S, et al. (2001) Limitations of conventional methods of self-monitoring of blood glucose: lessons learned from 3 days of continuous glucose sensing in pediatric patients with type 1 diabetes. *Diabetes Care* 24: 1858-1862.
 33. Leffingwell TR, Cooney NJ, Murphy JG, Luczak S, Rosen G, et al. (2013) Continuous objective monitoring of alcohol use: twenty-first century measurement using transdermal sensors. *Alcohol Clin Exp Res* 37: 16-22.
 34. Ritz T, Rosenfield D, Dewilde S, Steptoe A (2010) Daily mood, shortness of breath, and lung function in asthma: concurrent and prospective associations. *J Psychosom Res* 69: 341-351.
 35. Cramer JA, Mattson RH, Prevey ML, Scheyer RD, Ouellette VL (1989) How often is medication taken as prescribed: a novel assessment technique. *JAMA* 261: 3273-3277.
 36. Saito M, Kumano H, Yoshiuchi K, Kokubo N, Ohashi K, et al. (2005) Symptom profile of multiple chemical sensitivity in actual life. *Psychosom Med* 67: 318-325.
 37. Thiele C., Laireiter A-R., U. B (2002) Diaries in clinical psychology and psychotherapy: a selective review. *Clin Psychol Psychother* 9: 1-37.
 38. Vranceanu AM, Gallo LC, Bogart LM (2009) Depressive symptoms and momentary affect: the role of social interaction variables. *Depress Anxiety* 26: 464-470.
 39. Kim J, Kikuchi H, Yamamoto Y (2013) Systematic comparison between ecological momentary assessment and day reconstruction method for fatigue and mood states in healthy adults. *Br J Health Psychol* 18: 155-167.
 40. Stone AA, Smyth JM, Pickering T, Schwartz J (1996) Daily mood variability: form of diurnal

- patterns and determinants of diurnal patterns. *J Appl Soc Psychol* 26: 1286-1305.
41. Steptoe A, Leigh ES, Kumari M (2011) Positive affect and distressed affect over the day in older people. *Psychol Aging* 26: 956-965.
 42. Farquharson B, Bell C, Johnston D, Jones M, Schofield P, et al. (2013) Nursing stress and patient care: real-time investigation of the effect of nursing tasks and demands on psychological stress, physiological stress, and job performance: study protocol. *J Adv Nurs* 69: 2327-2335.
 43. Thewissen V, Bentall RP, Lecomte T, van Os J, Myin-Germeys I (2008) Fluctuations in self-esteem and paranoia in the context of daily life. *J Abnorm Psychol* 117: 143-153.
 44. Carels RA, Douglass OM, Cacciapaglia HM, O'Brien WH (2004) An ecological momentary assessment of relapse crises in dieting. *J Consult Clin Psychol* 72: 341-348.
 45. Dunton GF, Atienza AA, Castro CM, King AC (2009) Using ecological momentary assessment to examine antecedents and correlates of physical activity bouts in adults age 50+ years: a pilot study. *Ann Behav Med* 38: 249-255.
 46. Wichers M, Peeters F, Rutten BP, Jacobs N, Derom C, et al. (2012) A time-lagged momentary assessment study on daily life physical activity and affect. *Health Psychol* 31: 135-144.
 47. Shiffman S, Balabanis MH, Gwaltney CJ, Paty JA, Gnys M, et al. (2007) Prediction of lapse from associations between smoking and situational antecedents assessed by ecological momentary assessment. *Drug Alcohol Depend* 91: 159-168.
 48. Chandra S, Scharf D, Shiffman S (2011) Within-day temporal patterns of smoking, withdrawal symptoms, and craving. *Drug Alcohol Depend* 117: 118-125.
 49. Shiffman S, Kirchner TR (2009) Cigarette-by-cigarette satisfaction during ad libitum smoking. *J Abnorm Psychol* 118: 348-359.
 50. Hillbrand M, Waite BM (1994) The everyday experience of an institutionalized sex offender: an idiographic application of the experience sampling method. *Arch Sex Behav* 23: 453-463.
 51. Muller A, Mitchell JE, Crosby RD, Cao L, Johnson J, et al. (2012) Mood states preceding and following compulsive buying episodes: an ecological momentary assessment study. *Psychiatry Res* 200: 575-580.
 52. Helgeson VS, Lopez LC, Kamarck T (2009) Peer relationships and diabetes: retrospective and ecological momentary assessment approaches. *Health Psychol* 28: 273-282.
 53. Rutledge T, Stucky E, Dollarhide A, Shively M, Jain S, et al. (2009) A real-time assessment of work stress in physicians and nurses. *Health Psychol* 28: 194-200.
 54. Mulvaney SA, Rothman RL, Dietrich MS, Wallston KA, Grove E, et al. (2012) Using mobile phones to measure adolescent diabetes adherence. *Health Psychol* 31: 43-50.
 55. McCarthy DE, Piasecki TM, Lawrence DL, Jorenby DE, Shiffman S, et al. (2008) Psychological mediators of bupropion sustained-release treatment for smoking cessation. *Addiction* 103: 1521-1533.
 56. Kikuchi H, Yoshiuchi K, Yamamoto Y, Komaki G, Akabayashi A (2012) Diurnal variation of tension-type headache intensity and exacerbation: an investigation using computerized ecological momentary assessment. *Biopsychosoc Med* 6: 18.
 57. Everhart RS, Smyth JM, Santuzzi AM, Fiese BH (2010) Validation of the Asthma Quality of Life Questionnaire with momentary assessments of symptoms and functional limitations in patient daily life. *Respir Care* 55: 427-432.
 58. D'Amato G, Cecchi L (2008) Effects of climate change on environmental factors in respiratory allergic diseases. *Clin Exp Allergy* 38: 1264-1274.
 59. Henry JA, Galvez G, Turbin MB, Thielman EJ, McMillan GP, et al. (2012) Pilot study to evaluate ecological momentary assessment of tinnitus. *Ear Hear* 33: 179-290.

60. Epstein DH, Marrone GF, Heishman SJ, Schmittner J, Preston KL (2010) Tobacco, cocaine, and heroin: craving and use during daily life. *Addict Behav* 35: 318-324.
61. Freedman MJ, Lester KM, McNamara C, Milby JB, Schumacher JE (2006) Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. *J Subst Abuse Treat* 30: 105-111.
62. Weinland SR, Morris CB, Hu YM, Leserman J, Bangdiwala SI, et al. (2011) Characterization of episodes of irritable bowel syndrome using ecological momentary assessment. *Am J Gastroenterol* 106: 1813-1820.
63. Jern P, Gunst A, Sandqvist F, Sandnabba NK, Santtila P (2011) Using Ecological Momentary Assessment to investigate associations between ejaculatory latency and control in partnered and non-partnered sexual activities. *J Sex Res* 48: 316-324.
64. Munsch S, Meyer AH, Milenkovic N, Schlup B, Margraf J, et al. (2009) Ecological momentary assessment to evaluate cognitive-behavioral treatment for binge eating disorder. *Int J Eat Disord* 42: 648-657.
65. Zunker C, Peterson CB, Crosby RD, Cao L, Engel SG, et al. (2011) Ecological momentary assessment of bulimia nervosa: does dietary restriction predict binge eating? *Behav Res Ther* 49: 714-717.
66. Burd C, Mitchell JE, Crosby RD, Engel SG, Wonderlich SA, et al. (2009) An assessment of daily food intake in participants with anorexia nervosa in the natural environment. *Int J Eat Disord* 42: 371-374.
67. Rosen PJ, Epstein JN, Van Orden G (2013) I know it when I quantify it: ecological momentary assessment and recurrence quantification analysis of emotion dysregulation in children with ADHD. *Atten Defic Hyperact Disord* 5: 283-294.
68. Knouse LE, Mitchell JT, Brown LH, Silvia PJ, Kane MJ, et al. (2008) The expression of adult ADHD symptoms in daily life: an application of experience sampling methodology. *J Atten Disord* 11: 652-663.
69. Ebner-Priemer UW, Kuo J, Schlotz W, Kleindienst N, Rosenthal MZ, et al. (2008) Distress and affective dysregulation in patients with borderline personality disorder: a psychophysiological ambulatory monitoring study. *J Nerv Ment Dis* 196: 314-320.
70. Tan PZ, Forbes EE, Dahl RE, Ryan ND, Siegle GJ, et al. (2012) Emotional reactivity and regulation in anxious and nonanxious youth: a cell-phone ecological momentary assessment study. *J Child Psychol Psychiatry* 53: 197-206.
71. Pfaltz MC, Michael T, Grossman P, Margraf J, Wilhelm FH (2010) Instability of physical anxiety symptoms in daily life of patients with panic disorder and patients with posttraumatic stress disorder. *J Anxiety Disord* 24: 792-798.
72. Helbig-Lang S, Lang T, Petermann F, Hoyer J (2012) Anticipatory anxiety as a function of panic attacks and panic-related self-efficacy: an ambulatory assessment study in panic disorder. *Behav Cogn Psychother* 40: 590-604.
73. Forbes EE, Hariri AR, Martin SL, Silk JS, Moyses DL, et al. (2009) Altered striatal activation predicting real-world positive affect in adolescent major depressive disorder. *Am J Psychiatry* 166: 64-73.
74. Bylsma LM, Taylor-Clift A, Rottenberg J (2011) Emotional reactivity to daily events in major and minor depression. *J Abnorm Psychol* 120: 155-167.
75. Peeters F, Nicolson NA, Berkhof J, Delespaul P, deVries M (2003) Effects of daily events on mood states in major depressive disorder. *J Abnorm Psychol* 112: 203-211.
76. Kim J, Nakamura T, Kikuchi H, Sasaki T, Yamamoto Y (2013) Co-variation of depressive mood

- and locomotor dynamics evaluated by ecological momentary assessment in healthy humans. *PLoS ONE* 8: e74979.
77. Havermans R, Nicolson NA, Berkhof J, deVries MW (2010) Mood reactivity to daily events in patients with remitted bipolar disorder. *Psychiatry Res* 179: 47-52.
 78. Kimhy D, Delespaul P, Corcoran C, Ahn H, Yale S, et al. (2006) Computerized experience sampling method (ESMc): assessing feasibility and validity among individuals with schizophrenia. *J Psychiatr Res* 40: 221-230.
 79. Granholm E, Ben-Zeev D, Fulford D, Swendsen J (2013) Ecological Momentary Assessment of social functioning in schizophrenia: impact of performance appraisals and affect on social interactions. *Schizophr Res* 145: 120-124.
 80. Granholm E, Loh C, Swendsen J (2008) Feasibility and validity of computerized ecological momentary assessment in schizophrenia. *Schizophr Bull* 34: 507-514.
 81. Sorbi MJ, Mak SB, Houtveen JH, Kleiboer AM, van Doornen LJ (2007) Mobile Web-based monitoring and coaching: feasibility in chronic migraine. *J Med Internet Res* 9: e38.
 82. Jonasson G, Carlsen KH, Sodal A, Jonasson C, Mowinckel P (1999) Patient compliance in a clinical trial with inhaled budesonide in children with mild asthma. *Eur Respir J* 14: 150-154.
 83. Macdonell KE, Naar-King S, Murphy DA, Parsons JT, Huszti H (2011) Situational temptation for HIV medication adherence in high-risk youth. *AIDS Patient Care STDS* 25: 47-52.
 84. Basen-Engquist K, Carmack CL, Li Y, Brown J, Jhingran A, et al. (2013) Social-cognitive theory predictors of exercise behavior in endometrial cancer survivors. *Health Psychol*.
 85. Focht BC, Ewing V, Gauvin L, Rejeski WJ (2002) The unique and transient impact of acute exercise on pain perception in older, overweight, or obese adults with knee osteoarthritis. *Ann Behav Med* 24: 201-210.
 86. Conner TS, Barrett LF (2012) Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosom Med* 74: 327-337.
 87. Todd M, Armeli S, Tennen H, Carney MA, Ball SA, et al. (2005) Drinking to cope: a comparison of questionnaire and electronic diary reports. *J Stud Alcohol* 66: 121-129.
 88. Kamarck TW, Shiffman SM, Smithline L, Goodie JL, Paty JA, et al. (1998) Effects of task strain, social conflict, and emotional activation on ambulatory cardiovascular activity: daily life consequences of recurring stress in a multiethnic adult sample. *Health Psychol* 17: 17-29.
 89. Affleck G, Tennen H, Urrows S, Higgins P, Abeles M, et al. (1998) Fibromyalgia and women's pursuit of personal goals: a daily process analysis. *Health Psychol* 17: 40-47.
 90. Goldstein IB, Jamner LD, Shapiro D (1992) Ambulatory blood pressure and heart rate in healthy male paramedics during a workday and a nonworkday. *Health Psychol* 11: 48-54.
 91. Kamarck TW, Schwartz JE, Shiffman S, Muldoon MF, Sutton-Tyrrell K, et al. (2005) Psychosocial stress and cardiovascular risk: what is the role of daily experience? *J Pers* 73: 1749-1774.
 92. Kamarck TW, Muldoon MF, Shiffman SS, Sutton-Tyrrell K (2007) Experiences of demand and control during daily life are predictors of carotid atherosclerotic progression among healthy men. *Health Psychol* 26: 324-332.
 93. Shiffman S, Elash CA, Paton SM, Gwaltney CJ, Paty JA, et al. (2000) Comparative efficacy of 24-hour and 16-hour transdermal nicotine patches for relief of morning craving. *Addiction* 95: 1185-1195.
 94. Kikuchi H, Yoshiuchi K, Miyasaka N, Ohashi K, Yamamoto Y, et al. (2006) Reliability of recalled self-report on headache intensity: investigation using ecological momentary assessment technique. *Cephalalgia* 26: 1335-1343.

95. Reis HT, Wheeler L (1991) Studying social-interaction with the Rochester Interaction Record. *Adv Exp Soc Psychol* 24: 269-318.
96. Shiffman S, Gwaltney CJ, Balabanis MH, Liu KS, Paty JA, et al. (2002) Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment. *J Abnorm Psychol* 111: 531-545.
97. Muraven M, Collins RL, Shiffman S, Paty JA (2005) Daily fluctuations in self-control demands and alcohol intake. *Psychol Addict Behav* 19: 140-147.
98. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR (2002) Patient non-compliance with paper diaries. *BMJ* 324: 1193-1194.
99. Shiffman S (2009) Ecological momentary assessment (EMA) in studies of substance use. *Psychol Assess* 21: 486-497.
100. Litt MD, Cooney NL, Morse P (1998) Ecological Momentary Assessment (EMA) with treated alcoholics: methodological problems and potential solutions. *Health Psychol* 17: 48-52.
101. Shiffman S (2009) How many cigarettes did you smoke? Assessing cigarette consumption by global report, Time-Line Follow-Back, and ecological momentary assessment. *Health Psychol* 28: 519-526.
102. Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M (2002) Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers. *Psychol Addict Behav* 16: 205-211.
103. Cruise CE, Broderick J, Porter L, Kaell A, Stone AA (1996) Reactive effects of diary self-assessment in chronic pain patients. *Pain* 67: 253-258.
104. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, et al. (2003) Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. *Pain* 104: 343-351.
105. Singer JD, Willett JB (2003) *Applied longitudinal data analysis: modeling change and event occurrence*. Oxford: Oxford University Press.
106. Hox JJ, Roberts JK (2011) *Handbook of advanced multilevel analysis*. New York: Routledge.
107. Schwartz JE, Stone AA (1998) Strategies for analyzing ecological momentary assessment data. *Health Psychol* 17: 6-16.
108. Stone AA, Broderick JE, Porter LS, Kaell AT (1997) The experience of rheumatoid arthritis pain and fatigue: examining momentary reports and correlates over one week. *Arthritis Care Res* 10: 185-193.
109. Stone AA, Broderick JE, Shiffman SS, Schwartz JE (2004) Understanding recall of weekly pain from a momentary assessment perspective: absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain* 107: 61-69.
110. Friedberg F, Sohl SJ (2008) Memory for fatigue in chronic fatigue syndrome: the relation between weekly recall and momentary ratings. *Int J Behav Med* 15: 29-33.
111. Smith TW, Birmingham W, Uchino BN (2012) Evaluative threat and ambulatory blood pressure: cardiovascular effects of social stress in daily experience. *Health Psychol* 31: 763-766.
112. Schwarz N (1999) Self-reports - How the questions shape the answers. *Am Psychol* 54: 93-105.
113. Banthia R, Malcarne VL, Roesch SC, Ko CM, Greenbergs HL, et al. (2006) Correspondence between daily and weekly fatigue reports in breast cancer survivors. *J Behav Med* 29: 269-279.
114. Hacker ED, Ferrans CE (2007) Ecological momentary assessment of fatigue in patients receiving intensive cancer therapy. *J Pain Symptom Manage* 33: 267-275.
115. Jhamb M, Weisbord SD, Steel JL, Unruh M (2008) Fatigue in patients receiving maintenance

- dialysis: a review of definitions, measures, and contributing factors. *Am J Kidney Dis* 52: 353-365.
116. Sohl SJ, Friedberg F (2008) Memory for fatigue in chronic fatigue syndrome: relationships to fatigue variability, catastrophizing, and negative affect. *Behav Med* 34: 29-35.
 117. McColl E (2004) Best practice in symptom assessment: a review. *Gut* 53: 49-54.
 118. Kahneman D, Krueger AB, Schkade D, Schwarz N, Stone AA (2006) Would you be happier if you were richer? A focusing illusion. *Science* 312: 1908-1910.
 119. Dockray S, Grant N, Stone AA, Kahneman D, Wardle J, et al. (2010) A comparison of affect ratings obtained with ecological momentary assessment and the day reconstruction method. *Soc Indic Res* 99: 269-283.
 120. Fukui I (1997) The depression and anxiety mood scale (DAMS): scale development and validation. *Jpn J Behav Ther* 23: 83-93.
 121. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, et al. (2000) Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 32: S498-504.
 122. Krueger AB, Stone AA (2008) Assessment of pain: a community-based diary survey in the USA. *Lancet* 371: 1519-1525.
 123. Kumar S, Nilsen W, Pavel M, Srivastava M (2013) Mobile health: revolutionizing healthcare through transdisciplinary research. *IEEE Comput* 46: 28-35.
 124. American Psychiatric Association (2000) *Diagnostic and statistical manual of mental disorders*. Washington, D.C.: American Psychiatric Press.
 125. Jahng S, Solhan MB, Tomko RL, Wood PK, Piasecki TM, et al. (2011) Affect and alcohol use: an ecological momentary assessment study of outpatients with borderline personality disorder. *J Abnorm Psychol* 120: 572-584.
 126. Nakamura T, Kiyono K, Yoshiuchi K, Nakahara R, Struzik Z, et al. (2007) Universal scaling law in human behavioral organization. *Phys Rev Lett* 99: 138103.
 127. Nakamura T, Takumi T, Takano A, Aoyagi N, Yoshiuchi K, et al. (2008) Of mice and men--universality and breakdown of behavioral organization. *PLoS ONE* 3: e2050.
 128. Kikuchi H, Yoshiuchi K, Ohashi K, Yamamoto Y, Akabayashi A (2007) Tension-type headache and physical activity: an actigraphic study. *Cephalalgia* 27: 1236-1243.
 129. Sano W, Nakamura T, Yoshiuchi K, Kitajima T, Tsuchiya A, et al. (2012) Enhanced persistency of resting and active periods of locomotor activity in schizophrenia. *PLoS ONE* 7: e43539.
 130. Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29: 1165-1188.
 131. Suls J, Bunde J (2005) Anger, anxiety, and depression as risk factors for cardiovascular disease: the problems and implications of overlapping affective dispositions. *Psychol Bull* 131: 260-300.
 132. Bleil ME, Gianaros PJ, Jennings JR, Flory JD, Manuck SB (2008) Trait negative affect: toward an integrated model of understanding psychological risk for impairment in cardiac autonomic function. *Psychosom Med* 70: 328-337.
 133. Thieme K, Turk DC, Flor H (2004) Comorbid depression and anxiety in Fibromyalgia syndrome: relationship to somatic and psychosocial variables. *Psychosom Med* 66: 837-844.
 134. Demyttenaere K, De Fruyt J, Stahl SM (2005) The many faces of fatigue in major depressive disorder. *Int J Neuropsychopharmacol* 8: 93-105.
 135. Kennedy SH (2008) Core symptoms of major depressive disorder: relevance to diagnosis and treatment. *Dialogues Clin Neurosci* 10: 271-277.

136. Tylee A, Gastpar M, Lepine JP, Mendlewicz J (1999) DEPRES II (Depression Research in European Society II): a patient survey of the symptoms, disability and current management of depression in the community. DEPRES Steering Committee. *Int Clin Psychopharmacol* 14: 139-151.
137. Baldwin DS, Papakostas GI (2006) Symptoms of fatigue and sleepiness in major depressive disorder. *J Clin Psychiatry* 67 Suppl 6: 9-15.
138. Skapinakis P, Lewis G, Mavreas V (2003) Unexplained fatigue syndromes in a multinational primary care sample: specificity of definition and prevalence and distinctiveness from depression and generalized anxiety. *Am J Psychiatry* 160: 785-787.
139. Tchekmedyian NS, Kallich J, McDermott A, Fayers P, Erder MH (2003) The relationship between psychologic distress and cancer-related fatigue. *Cancer* 98: 198-203.
140. Linhart H (1988) A test whether two AIC's differ significantly. *S Afr Stat J* 22: 153-161.
141. Stavrakakis N, de Jonge P, Ormel J, Oldehinkel AJ (2012) Bidirectional prospective associations between physical activity and depressive symptoms. The TRAILS Study. *J Adolesc Health Care* 50: 503-508.
142. Lindwall M, Larsman P, Hagger MS (2011) The reciprocal relationship between physical activity and depression in older European adults: a prospective cross-lagged panel design using SHARE data. *Health Psychol* 30: 453-462.
143. Harris AH, Cronkite R, Moos R (2006) Physical activity, exercise coping, and depression in a 10-year cohort study of depressed patients. *J Affect Disord* 93: 79-85.
144. Strawbridge WJ (2002) Physical activity reduces the risk of subsequent depression for older adults. *Am J Epidemiol* 156: 328-334.
145. Camacho TC, Roberts RE, Lazarus NB, Kaplan GA, Cohen RD (1991) Physical activity and depression: evidence from the Alameda County Study. *Am J Epidemiol* 134: 220-231.
146. Damasio AR (2004) *Looking for Spinoza: joy, sorrow, and the feeling brain*. London: Vintage.
147. Harvey SB, Hotopf M, Overland S, Mykletun A (2010) Physical activity and common mental disorders. *Br J Psychiatry* 197: 357-364.
148. Goodwin RD (2003) Association between physical activity and mental disorders among adults in the United States. *Prev Med* 36: 698-703.
149. Faurholt-Jepsen M, Brage S, Vinberg M, Christensen EM, Knorr U, et al. (2012) Differences in psychomotor activity in patients suffering from unipolar and bipolar affective disorder in the remitted or mild/moderate depressive state. *J Affect Disord* 141: 457-463.
150. St-Amand J, Provencher MD, Belanger L, Morin CM (2013) Sleep disturbances in bipolar disorder during remission. *J Affect Disord* 146: 112-119.
151. Joffe H, Soares CN, Thurston RC, White DP, Cohen LS, et al. (2009) Depression is associated with worse objectively and subjectively measured sleep, but not more frequent awakenings, in women with vasomotor symptoms. *Menopause* 16: 671-679.
152. Robillard R, Naismith SL, Rogers NL, Ip TK, Hermens DF, et al. (2013) Delayed sleep phase in young people with unipolar or bipolar affective disorders. *J Affect Disord* 145: 260-263.
153. Baune BT, Caliskan S, Todder D (2007) Effects of adjunctive antidepressant therapy with quetiapine on clinical outcome, quality of sleep and daytime motor activity in patients with treatment-resistant depression. *Hum Psychopharmacol* 22: 1-9.
154. Hamilton M (1960) A rating scale for depression. *J Neurol Neurosurg Psychiatry* 23: 56-62.
155. Beck AT, Steer RA, Brown GK (1996) *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
156. Kojima M, Furukawa TA, Takahashi H, Kawai M, Nagaya T, et al. (2002) Cross-cultural

- validation of the Beck Depression Inventory-II in Japan. *Psychiatry Res* 110: 291-299.
157. Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38: 963-974.
158. Zeger SL, Liang KY (1992) An overview of methods for the analysis of longitudinal data. *Stat Med* 11: 1825-1839.
159. Barabasi AL (2005) The origin of bursts and heavy tails in human dynamics. *Nature* 435: 207-211.
160. Proekt A, Banavar JR, Maritan A, Pfaff DW (2012) Scale invariance in the dynamics of spontaneous behavior. *Proc Natl Acad Sci U S A* 109: 10564-10569.
161. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439: 462-465.
162. Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. *Nature* 453: 779-782.
163. Malmgren RD, Stouffer DB, Motter AE, Amaral LA (2008) A Poissonian explanation for heavy tails in e-mail communication. *Proc Natl Acad Sci U S A* 105: 18153-18158.
164. Nakamura T, Takumi T, Takano A, Hatanaka F, Yamamoto Y (2013) Characterization and modeling of intermittent locomotor dynamics in clock gene-deficient mice. *PLoS ONE* 8: e58884.
165. *PSYCHE: Personalised monitoring SYstems for Care in mental HEalth*. Retrieved April 2, 2013 from <http://www.psyche-project.org/>
166. *MONARCA - MONitoring, treAtment and pRediCtion of bipolar Disorder Episodes*. Retrieved April 2, 2013 from <http://www.monarca-project.eu/>
167. Heron KE, Smyth JM (2010) Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *Br J Health Psychol* 15: 1-39.
168. Struzik ZR, Yoshiuchi K, Sone M, Ishikawa T, Kikuchi H, et al. (2007) "Mobile nurse" platform for ubiquitous medicine. *Methods Inf Med* 46: 130-134.
169. Vapnik VN (1995) *The nature of statistical learning theory*. New York: Springer.
170. Bishop CM (1995) *Neural networks for pattern recognition*. Oxford: Clarendon Press.
171. Williams DA, Gendreau M, Hufford MR, Groner K, Gracely RH, et al. (2004) Pain assessment in patients with fibromyalgia syndrome: a consideration of methods for clinical trials. *Clin J Pain* 20: 348-356.
172. Redelmeier DA, Katz J, Kahneman D (2003) Memories of colonoscopy: a randomized trial. *Pain* 104: 187-194.

SUPPLEMENTARY INFORMATION

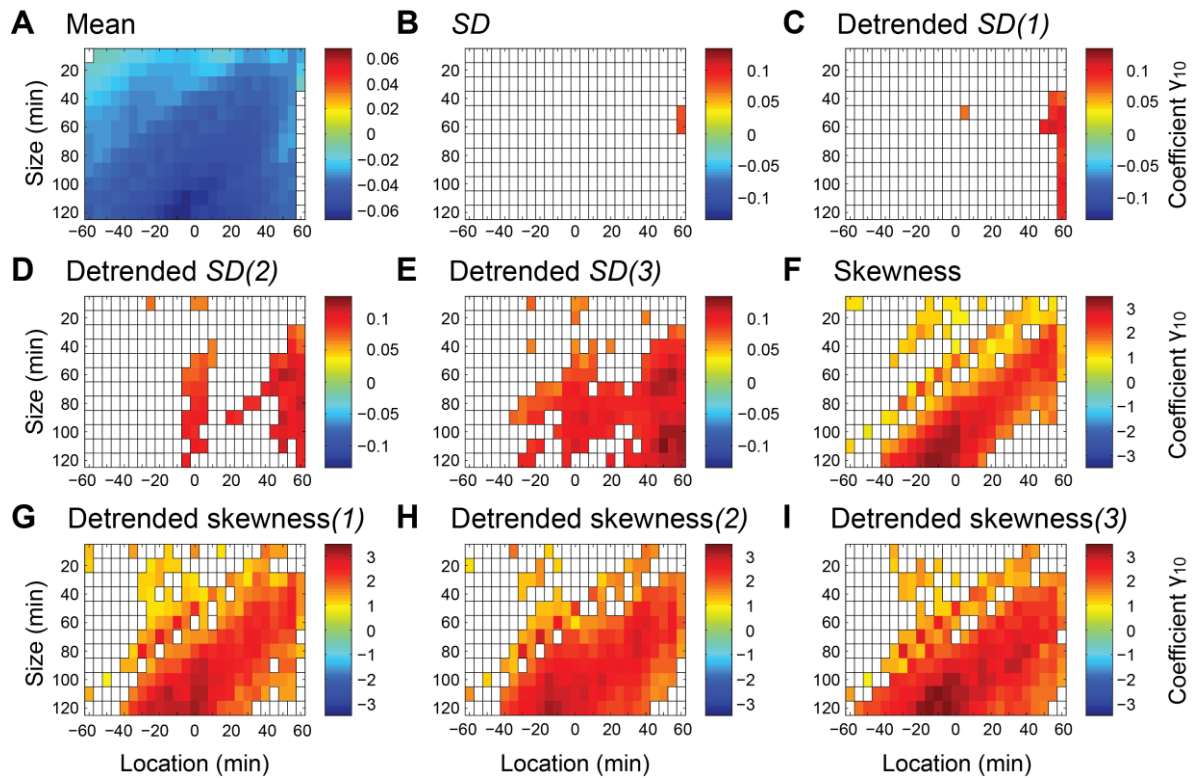
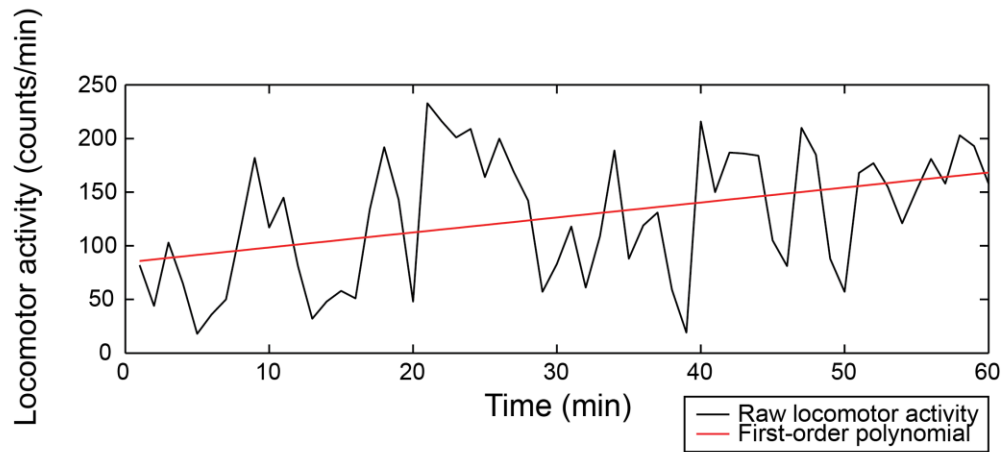


Figure S1. Dependency of univariate model coefficients for depressive mood on different time frames. The estimated values of the univariate model coefficient for depressive mood scores are shown in a colored matrix form consisting of 25 rows (different locations) and 12 columns (different sizes). Each grid cell indicates some specific location and size of a time frame used for calculating the local statistics of locomotor activity. A color in each cell represents the value of the model coefficient (γ_{10}) of the predictors: (A) Mean, (B) SD , (C) Detrended $SD(1)$, (D) Detrended $SD(2)$, (E) Detrended $SD(3)$, (F) Skewness, (G) Detrended skewness(1), (H) Detrended skewness(2), and (I) Detrended skewness(3). Note that these statistics with m -th order polynomial detrending are denoted as detrended $SD(m)$ and detrended skewness(m). False discovery rate with a q value of 0.05 was used as the multiple comparison adjustment. Only significant cases are shown by colors. Color bars indicate the value of the model coefficient. The univariate model used for the analysis is as follows; Depressive mood score $_{ij} = \gamma_{00} + \gamma_{10}$ (Local statistics of locomotor activity $_{ij}$) + $\zeta_{0i} + \epsilon_{ij}$ (see 1.4.1 for details).

Text S1. Detrending procedure for calculating detrended *SD* and skewness.

- 1) A polynomial function was fitted to locomotor activity data within a time frame to estimate a “trend” of the data. In this study, I considered the cases up to the third-order polynomials (Figure S1).
- 2) The estimated trend was removed by subtracting from the original data to derive the detrended locomotor activity data (Figure S2).
- 3) From the detrended data, *SD* and skewness were calculated.

A Raw locomotor activity



B Detrended locomotor activity

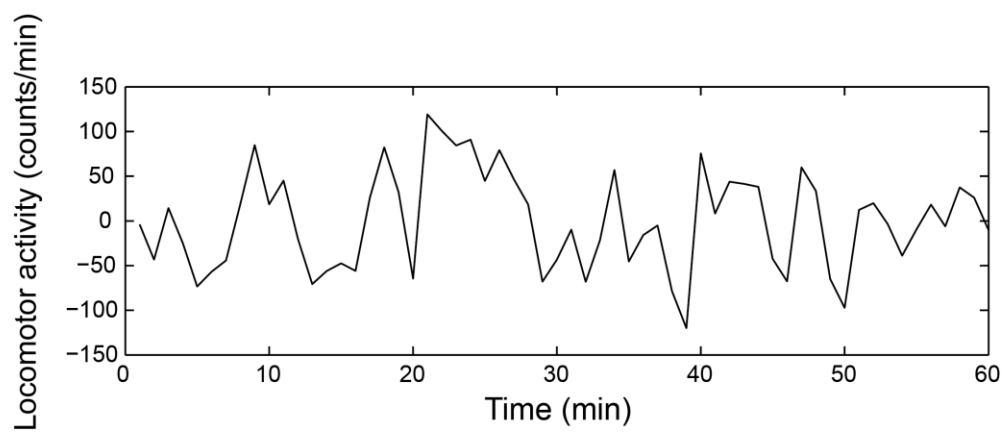


Figure S2. Detrending procedure for calculating detrended *SD* and skewness. (A): A raw locomotor activity time series for 60 min and the best fitted first-order polynomial line (red). (B): The detrended locomotor activity derived by subtracting the fitted line for the original data.