5-26-2006

# An ecological momentary assessment of retrospective memory accuracy in patients with obsessive-compulsive disorder

Andrew T. Gloster

AN ECOLOGICAL MOMENTARY ASSESSMENT OF RETROSPECTIVE MEMORY

ACCURACY IN PATIENTS WITH OBSESSIVE-COMPULSIVE DISORDER

by

Andrew T. Gloster

Dissertation

Submitted to the Department of Psychology

Eastern Michigan University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Clinical Psychology

Dissertation Committee:

David C. S. Richard, Ph.D., Chair

Joseph Himle, Ph.D.

John Knapp, Ph.D.

Ellen Koch, Ph.D.

James Thornton, Ph.D.

May 26, 2006

Ypsilanti, Michigan

Acknowledgments

This study represents the culmination of my academic endeavors to date. I could not have arrived at this point without various forms of assistance en route. I have been blessed with marvelous mentors and dear friends; I hereby wish to formally acknowledge these people and express my gratitude. No matter where life's road may take me, you are in my thoughts. My successes are your successes.

Dr. David C. S. Richard served as my dissertation chair, academic advisor, and mentor par excellence. In this role, he continually challenged me and helped me grow. Likewise, he trusted me and afforded me valuable opportunities to help launch my career. My writing skills have improved exponentially under his tutelage. Without his patient and steadfast editing, this manuscript would be twice as long. Beyond our formal relationship, Dr. Richard has become a friend. I look forward to the next stages of our friendship and future collaborations.

Dr. James Thornton offered statistical consulting second to none. Dr. Thornton was tireless in his desire to help peel away the mysteries of advanced statistics. Aside from allowing me to audit his course, Dr. Thornton spent many hours consulting and teaching, always demonstrating enthusiasm regardless of whether he was in his office, at home on sabbatical, on the way to the store, or on a Florida beach. I look forward to our future collaborations and continued friendship.

Dr. Joe Himle served various important roles in my training and in this dissertation. First, he opened the doors of the University of Michigan Anxiety Disorders Clinic to me offering training and research opportunities. More importantly, he opened the doors of his office, allowing for stimulating exchanges of ideas. His leadership style never failed to inspire, and I feel lucky to have been a part of this. I look forward to future collaborations and continued friendship.

Dr. Ellen Koch was my clinical supervisor while at the University of Michigan Anxiety Disorders Clinic and at Eastern Michigan University. Dr. Koch's enthusiasm for anxiety treatment and research is inspiring. She never grew tired of my questions and supported all of my endeavors in and out of the therapy room in ways too numerous to list. I credit Dr. Koch with assisting me in refining my clinical skills and helping navigate the important exchange between clinical ideas and meaningful research. I look forward to future collaborations and continued friendship.

Dr. John Knapp has served on both my thesis and dissertation committees. Throughout the span of these two projects, he has helped me sharpen my research and statistical skills. His fine eye for detail has helped ensure quality products, for which I am very grateful. Furthermore, Dr. Knapp has carried the weight of the department on his shoulders through tough times and never asked for praise. His leadership skills are to be emulated. I hope I can be half as good. Again, I look forward to future collaborations and continued friendship.

The completion of this project would have been impossible without the outstanding help I received from Heather Anson, Lisa O'Donnell, Elizabeth Yaelingh-Scoffins, and Elizabeth Nelson. To label these people merely the assessors of this project would downplay their roles as colleagues and friends. Their dedication to precision and excellence stand on their own and contributed heavily to the quality of the final project. However, answering my phone calls at nearly every possible time of day, exchanging ideas, cheer-leading, and counseling speak to their quality as friends.

This project would also have been impossible without the generosity of the University of Michigan Anxiety Disorders Clinic. First, I wish to thank them for taking a chance with an unknown practicum student. Their dedication to excellence in therapy, research, and teaching will guide me in my career. I specifically wish to thank my other clinical supervisors, Dr. Heather O'Mahen and Dr. Shelly Van Etten Lee, for challenging me to conceptualize at ever

higher levels. Dr. James Abelson has cultivated an atmosphere in the clinic of true cross-disciplinary collegiality. The results speak for themselves. Last, but not least, I could not have organized the entire project (or even a day for that matter) without the generous support of Laura Lokers, Susan Miller, and Jody Porter. Whether navigating the waters of the IRB recruiting participants, or simply lending an ear, you never failed to point me in the right direction with a smile on your face. Finally, I wish to thank Pam Schweitzer, Claudia Miller, and the rest of the clinicians who helped refer patients to the project.

Technology played a central role in this study. Without my technical and programming advisors, Ryan Sexton and Randy Manthy, this project would have remained simply a good idea. Likewise, I wish to thank Pendragon cooperation for their support during the programming and data collection phases of the study. Further, I wish to acknowledge Shawn Mason for his willingness to help out in various ways during the past year to help support data collection while I was on the other side of the country.

I further owe a debt of gratitude to the "Houston contingency." The newly ordained Dr. Cindy Kraus was an excellent dissertation partner and helped make many of those long study weekends more fun. Drs. Howard Rhoades and Charles Green were both excellent consultants and always willing to answer my statistical questions. I have learned a lot. Debra Liles came through in a pinch with little notice to help me learn a new statistical program. Finally, Drs. Patricia Averill, Melinda Stanley, and Angela Stotts have been excellent mentors this year and helped me make the transition to a professional.

Less directly but equally important are contributions from my previous mentors. Mr. Irving Sarin took a boy under his wing and showed him the breadth of the horizon. Thank you. Dr. Joel Sheveloff inspired and demonstrated excellence in teaching and mentoring. I hope I can live up to your standards. Dr. Harry Schröder took a chance with a musician and opened the doors to psychology in psychology's birthplace. Whatever I accomplish in this field will be a direct result of your support. Dr. Dean Lauterbach took on the thankless task of washing the green from behind my ears when I first entered graduate school. Thank you.

My family and friends have aided and supported me in a thousand different ways I could list, and likely another thousand I am unaware of. I could not have gotten here without your unyielding support and encouragement. Above all else, I thank all of you for continuing to support me despite our lean schedule of contact.

Finally, and most of all, I acknowledge my best friend and partner. Ulrike, you have waited for me for five years as I pursued this goal. You supported me in all my decisions. You helped me get up when I was down. You managed to watch out for me from the other side of the world. And you did all of this while making massive sacrifices. I can not thank you enough for your support. Ich liebe dich and look forward to starting my life with you anew. I can not imagine a better soul-mate.

Abstract

Numerous studies have demonstrated the pervasiveness of inaccuracies in patients' retrospective recall of their symptoms (e.g., Stone, Broderick, Shiffman, & Schwartz, 2004). Assessment methods that rely heavily on retrospective recall may lead to faulty clinical inferences should a patient's recall be biased or inaccurate. Despite lingering concerns about the accuracy of retrospective recall in a variety of clinical and nonclinical populations, investigators have not studied individuals diagnosed with obsessive-compulsive disorder (OCD). This is troubling given findings from laboratory studies that OCD patients may have deficits in episodic memory (Muller & Roberts, 2005). This study investigated memory accuracy in OCD patients using an ecological momentary assessment (EMA) research methodology. By using handheld computers to collect self-monitoring data in real time, EMA data served as a criterion against which retrospective recall was tested for accuracy.

Thirty-five patients diagnosed with OCD used a handheld computer to rate presence of OCD and related symptoms four times per day for a week. Patients estimated the frequency and duration of their behavior during the EMA self-monitoring phase. Results indicated that contrary to *a priori* hypotheses, OCD patients' retrospective recall of their EMA recorded symptoms were relatively accurate. Consistent with hypotheses and previous studies, reactivity to the EMA data collection procedure was not observed. Finally, the results suggest that despite participants' accuracy when recalling frequency and duration of symptoms, participants were inaccurate in estimating symptom covariance with supplemental items that measured non-OCD functioning (e.g., amount of sleep, current level of stress, etc.).

TABLE OF CONTENTS

List of Tables

List of Figures

AN ECOLOGICAL MOMENTARY ASSESSMENT OF RETROSPECTIVE MEMORY

ACCURACY IN PATIENTS WITH OBSESSIVE-COMPULSIVE DISORDER

Introduction

Clinical assessment relies on a patient's retrospective recall to estimate the severity,

frequency, and intensity of presenting problems and variables that moderate and mediate

problem expression. Although clinicians implicitly assume patients are reasonably accurate

when reporting behavioral dimensions of their problems, researchers have noted inaccuracies

in retrospective reports. Researchers have investigated the accuracy of individuals'

retrospective recall[1] across numerous behaviors and disorders. For example, retrospective

estimates of pain intensity (Stone et al., 2004), frequency of eating behaviors (Stein & Corte,

2003), coping behaviors (Stone et al., 1998), panic attacks (de Beurs, Lange, & Van Dyck,

1992), smoking lapses (Shiffman et al., 1997), and anxious cognitions (Marks & Hemsley,

1999), are systematically biased when compared to data obtained *in vivo*. As can be seen in

Table 1, studies that have investigated accuracy of retrospective recall vary with respect to

sample size, the duration of the targeted time frame, and behavioral dimensions assessed

(e.g., frequency, intensity, etc.). Nevertheless, all of these studies found at least some

evidence of inaccuracy in retrospective recall, though the direction (i.e., overestimation vs.

underestimation) of the inaccuracies differed across studies.  However, early studies

investigating smoking frequency (Frederiksen, Epstein, & Kosevsky, 1975) and general

mood (Parkinson, Briner, Reynolds, & Totterdell, 1995) failed to find discrepancies in

patients' retrospective estimates of their own behavior and data collected *in vivo*. Thus,

although not universal, retrospectively obtained data may be inaccurate (i.e., random error),

biased (i.e., systematic error), or both (Shiffman et al., 1997). This study examined how accurate patients with Obsessive-Compulsive Disorder (OCD) were when recalling their own behavior.

*Factors affecting accuracy of retrospective recall*

One factor known to affect retrospective recall accuracy is the way in which the person recounts prior events. When recalling instances of their own behavior, people either estimate the frequency of a given behavior or enumerate individual instances of its occurrence (Sudman, Bradburn, & Schwarz, 1996). Estimation strategies appear to be used more often with frequently occurring behavior, for events further in the past, and as a result of a question's wording. Enumeration strategies are used with low frequency behaviors, for events in the recent past, and when prompts explicitly instruct individuals to use enumeration strategies (Menon & Yorkston, 2000). Furthermore, use of enumeration versus estimation strategies is a function of the temporal regularity of the behavior. That is, memories of temporally regular behaviors are consolidated and judgments are made by estimating mean frequencies of past behavior. Although estimation strategies can lead to accurate recall, high frequency behavior is difficult to estimate with any precision (Menon & Yorkston, 2000). Therefore, it may be the case that recall of behaviors that occur with great frequency is less accurate than recall of low frequency behaviors.

The level of current behavior has also been shown to influence accuracy in recall estimates of past behavior. For example, in a study of 106 undergraduates, Conway and Ross (1984) asked students to recall self-ratings of proficiency in several study skills they provided prior to participating in a study skills improvement program. Students in the improvement program recalled their previously evaluated study skills as significantly worse

---

[1] A glossary of technical terms used in this study can be found in Appendix D.

than students in the wait-list condition. In a recent review of errors in retrospective recall estimates, Tourangeau (2000) concluded that retrospectively recalled pre-treatment skills corresponded higher with current behavioral levels than with actual pre-treatment functioning. In addition, it may be that accuracy of retrospective recall is state-dependent in that negative affect potentiates recall of negative life events (Teasdale & Fogarty, 1979). Thus, the accuracy of retrospective recall may be a function of both the current level of behavior and the person's affect. As a result, patients may provide self-reports that underestimate behavioral variability, imply greater trait-like properties than is actually the case, and negatively influence the validity of a clinician's inferences about covariation of symptoms with changes in extrinsic variables (e.g., other behaviors, environments, contexts, mood states, etc.).

Theories of autobiographical memory (i.e., memory for one's own past thoughts and behavior) attempt to explain the accuracy of retrospective recall while simultaneously providing predictions about factors that mediate accurate recall. Brewer (1994) proposed that autobiographical memory is partially reconstructed. That is, personal memories are comprised both of aspects of the original experience and elements reconstructed from intervening factors such as repetition of the behavior. Memory consolidation processes thus impede accurate recall of events because information regarding certain key behavioral dimensions are lost (e.g., frequency, duration, etc.). Furthermore, Brewer states that high frequency events are cognitively organized by topographical similarity into self-schemas. A self-schema is a unit of related information from which general abstractions are made. For example, the self-schema for a person's food preference would include information concerning past experience with and impressions of different kinds of food. Information

related to these experiences is stored in the self-schema and a statement about food preference (e.g., "I like Italian food") reflects an abstraction drawn from information in the self-schema and not memories of individual episodes of behavior. Studies examining behaviors that follow predictable if slightly varied patterns (i.e., script-like behaviors such as birthdays, holidays, etc.) lend support to this position. For example, when asked to recall attributes such as location, time, social context, and feelings of script-like behaviors, students were more likely to utilize inferential retrieval processes than directly retrieving attributes of events. Inferential retrieval processes refer to constructing memories as a result of effortful processes whereas direct retrieval refers to memories generated immediately and automatically (Herman, 1994). When prompted to recall specific episodes or periods of stereotypical everyday experiences, or script-like behaviors (e.g., the previous week in therapy), individuals exhibiting repetitive stereotypic behaviors such as obsessions and compulsions may find the task difficult or impossible due to memory consolidation in self-schema formation. This may lead to the use of recall estimation strategies that yield inaccurate and/or biased autobiographical data.

Accuracy of retrospective recall has been investigated primarily with regard to overt behaviors (e.g., smoking, drinking, purging, etc.). The relative accuracy of retrospectively obtained reports of latent processes (e.g., thoughts, emotions, moods, etc.) is less precisely understood. This is troubling given that many theories of anxiety and mood disorders posit a central role for cognition as a causal or maintaining factor (e.g., Beck & Emery, 1985). Marks and Hemsley (1999) examined reports of physical and cognitive symptoms in patients with agoraphobia who completed an *in vivo* exposure task. Patients completed questionnaires in fear-provoking situations (e.g., supermarket, restaurant, busy street) and within 24 hours of

returning home. *In vivo* questions queried about the presence or absence of cognitive and physiological symptoms since beginning the exposure. The questionnaire completed at home included items querying about the presence or absence of symptoms in their feared situation in general. Results from the recall task showed that participants significantly overestimated the presence of some *in vivo* cognitions (e.g., belief that they were going to be sick, faint, and choke to death) but not for any physical symptoms. Thus, recall bias was evident in agoraphobic patients when retrospectively recalling frequency of anxiety-related cognitions.

*Obsessive Compulsive Disorder and Memory*

Obsessive-Compulsive Disorder (OCD), a debilitating anxiety disorder, is defined by repetitive, intrusive, undesired thoughts (obsessions) and/or repetitive, intentional behaviors (compulsions) that function to neutralize anxiety or distress (American Psychiatric Association, 2000; Karno, Golding, Sorenson, & Burnam, 1988). The behaviors must cause marked distress, be present for at least one hour per day, or significantly interfere with functioning.

Because accuracy of retrospective recall degrades as behavior increases in frequency (Menon & Yorkston, 2000), the implications for assessment and treatment of OCD are significant due to the high frequency of maladaptive cognitions and behavior in this population. As a result, OCD patients may be more susceptible to inaccurate or biased memory estimates of their symptoms than patients with other psychiatric disorders. For instance, a diagnostic requirement of OCD stipulates that target behaviors occur at least one hour per day. In contrast, the diagnostic criteria of Bulimia-Nervosa indicate that behaviors must occur minimally twice per week (American Psychiatric Association, 2000). Thus, all other things being equal, OCD patients should demonstrate more memory inaccuracies than

Bulimia-Nervosa patients as a result of the higher frequency behaviors. Although correspondence between retrospective report and data obtained *in vivo* has been investigated with lower frequency problem behaviors (e.g., agoraphobia, Marks & Hemsley, 1999; eating disorders, Stein & Corte, 2003), high frequency behaviors such as those present in OCD remain unexamined.

The accuracy of retrospective recall is also a function of the regularity with which a behavior occurs (Menon & Yorkston, 2000). Menon (1994) suggested that behaviors that occur with a fixed periodicity are stored in people's memories such that when asked to recall the frequency of these behaviors, individuals access the stored rate of behavior and apply it to the time frame under examination. Menon also suggested that the greater the stereotypical similarity in behavior over multiple occurrences, the greater the likelihood the behavior will be represented in memory as an abstraction rather than as separate episodic memories. By way of extension, OCD symptoms are presumably relatively similar across occurrences and may be relatively regular in periodicity. Both of these attributes suggest that individuals with OCD are more likely to use an estimation strategy than an enumeration strategy when recalling OCD behaviors.

The accuracy of retrospective recall is also complicated by secondary memory impairments that can occur in mood and anxiety-disordered patients. Attentional and interpretive biases are well established with anxious individuals (Barlow, 2002). Assuming that the accuracy of autobiographical memories is dependent initially on the quality of encoding processes, retrospective recall in anxiety disorders may be a function of impaired cognitive processing during both the storage and retrieval phases. In laboratory experiments with OCD patients, deficits have been observed in memory retrieval (Savage et al., 1996) and

implicit learning, or the ability to acquire knowledge through repetition but without conscious effort (Deckersbach et al., 2002). These findings suggest the possibility of structural brain abnormalities in patients with OCD. Specifically, corticostriatal system abnormalities in the limbic system appear to be related to failings in implicit learning. Thus, observed memory deficits may result from encoding difficulties.

Laboratory experiments suggest that relative to control subjects, OCD patients exhibit deficits across various memory tasks. Laboratory examination of memory deficits in patients diagnosed with OCD have concentrated on episodic memories, or personal events dated in the past, and concentrated on verbal and nonverbal stimuli (Muller & Roberts, 2005). In their excellent review, Muller and Roberts concluded that whereas the evidence for deficits in verbal memory is mixed, researchers have more consistently observed deficits in non-verbal memory relative to control groups. It is important to note, however, that studies of verbal and nonverbal memory utilize stimuli conducive for the laboratory (i.e., Wechsler Memory Scale). Although laboratory stimuli are important, empirical studies are necessary to test the generalizability and external validity of findings based on these stimuli.

One study has broadened the breadth of stimuli used in laboratory experiments to include personal experiences (Wilhelm, McNally, Baer, & Florin, 1997). Through the use of personally relevant stimuli, Wilhelm et al. found that memory biases in patients with OCD also manifest themselves in the form of overgenerality effects, or the tendency to recall categories of events when asked to recall specific instances. For example, when asked to tell the "...first *specific* personal memory that comes to mind" (Wilhelm et al., 1997, pg. 24) in response to a happy cue word, an overgeneral memory would be "when I watch movies" in comparison to the more specific memory "the night we went to the new restaurant."

Compared to healthy controls, OCD patients recalled a significantly lower percentage of specific memories and showed longer retrieval latencies during recall tasks. However, OCD patients with comorbid depression recalled statistically significantly fewer specific memories than OCD patients without depression. Wilhelm et al. concluded that depression and not OCD per se may mediate the overgenerality effect. Thus, Wilhelm et al. is noteworthy for two reasons. First, they expanded the range of stimuli used to study memory deficits in OCD patients to include personally relevant events. Second, they observed that the relatively inferior performance compared to control subjects may be specific to OCD patients or may be confounded by the presence of a mood disorder.

*Effects of memory bias on patient estimates of functional relations*

Clinician inferences based on inaccurate patient recall have the potential to negatively influence treatment planning and compromise treatment efficacy (Haynes, Leisen, & Blaine, 1997). Studies investigating the accuracy of autobiographical memories typically examine frequency counts or dichotomous endorsements of events. Clinical practice, however, normally requires patients to judge the covariance of symptoms, mood, and environmental stimuli. Unfortunately, only one published study has systematically examined the accuracy of symptom covariation estimates (O'Brien, 1995). In this study, advanced clinical psychology graduate students were presented hypothetical self-monitoring data of a patient's target behaviors (e.g., headache frequencies, durations, and intensities) and hypothesized controlling factors (e.g., stress level, number of arguments, sleep duration, and number of pain killers taken). Students in this experiment overinflated estimations of weak relationships and underinflated estimations of strong relationships (O'Brien).

Given these findings, it is likely that patients also have difficulty identifying functional relations. The term "functional relation" refers to whether two variables co-vary. The term does not imply causality, although a causal relationship is one kind of functional relation (Richard & Haynes, 2002). To date, no studies have examined how accurate individuals with a mood or anxiety disorder are in estimating functional relations. One would expect graduate students to be better at estimating the magnitude of co-varying relationships than individuals with a diagnosed mental disorder. However, it remains an untested empirical question if patients' retrospective estimations of functional relations yield the same inaccuracies noted in trained graduate students, if the inaccuracies are compounded due to memory biases, or if personally relevant data (i.e., from one's own life) mediate the effect.

*Ecological Momentary Assessment (EMA)*

Laboratory studies examining memory biases utilize highly controlled experiments with narrow ranges of stimuli (e.g., Stroop task, visual field manipulations, auditory channels, lexical tasks, etc.). These stimuli and experiments have gone a long way toward elucidating the vulnerabilities and phenomenology of patients with anxiety. Nevertheless, the generalizability of such findings requires empirical validation. Studies designed to examine the ecological validity of results discovered in laboratory studies will help specify the robustness of currently understood phenomena and illuminate practical implications.

Ecological Momentary Assessment (EMA) is a form of self-monitoring in which behaviors are recorded by individuals in their natural environment (Stone & Shiffman, 1994). In EMA studies, recordings are often made using a personal digital assistant, or PDA, that prompts the individual at regular intervals for a response. In a typical EMA study, patients

estimate the frequency of a behavior within a given time frame and record the estimate using the PDA. Recording may occur following a machine-generated beep (signal-contingent recording) or be self-initiated following the occurrence of the behavior (event-contingent recording). In either case, EMA studies minimize latencies between the occurrence and recording of behavior. In so doing, EMA represents a unique, ecologically valid opportunity to assess the accuracy of retrospective reports. Individuals therefore generate an incontrovertible index, or criterion, against which recall accuracy may be assessed.

Only one study has used EMA to examine OCD. Herman and Koran (1998) compared EMA data with clinician-rated symptom severity obtained prior to EMA data collection. Clinicians overestimated patients' frequency and intensity of OCD symptoms based on a Yale-Brown Obsessive-Compulsive Scale (see description below) interview. The discrepancy between clinician and EMA ratings was in the magnitude of $d = .95$ for obsessions and $d = .88$ for compulsions. Unfortunately, the design did not address accuracy of patients' retrospective recall because participants did not engage in autobiographical recall tasks. In contrast, the current study's design allowed for direct comparison of an individual's autobiographical recall against a criterion record of behavior, thus providing an assessment of retrospective recall accuracy in patients with OCD.

*Reactivity of EMA*

Despite potential advantages of EMA to increase accuracy of self-report, interpretation of the data must attend to issues of reactivity. Reactivity refers to change in one or more dimensions of behavior as a function of the method by which data are collected (Barton, Blanchard, & Veazy, 1999). Numerous factors are thought to increase reactivity to self-monitoring assessment methods. These include the valence of the target behavior

(positively valenced behaviors increase reactivity and negatively valenced behaviors decrease reactivity), motivation to change the behavior, visibility of the behavior (i.e., overt vs. covert), frequency of recording, timing of the recording (i.e., immediately before the behavior), and obtrusiveness of the monitoring device (Korotitsch & Nelson-Gray, 1999). With respect to EMA, recent studies have not shown changes in frequency of behaviors or motivation to change as a function of the monitoring method (Cruise, Broderick, Porter, Kaell, & Stone, 1996; Hufford, Shields, Shiffman, Paty, & Balabanis, 2002; Stein & Corte, 2003).

*Purpose*

This study examined whether memory inaccuracy and/or bias were present in the retrospective recall of OCD patients. Self-monitoring data collected via a handheld computer was used as the criterion against which the accuracy of retrospective recall was assessed. Detection of inaccurate autobiographical memories has important implications for clinical judgment. Namely, OCD patients may be inaccurate in retrospectively reporting the frequency and/or intensity of their symptoms. Given that case conceptualizations depend upon the accuracy of patient self-report, identification and quantification of inaccurate recall is important. Further, if large effect sizes exist between retrospectively recalled data and EMA criterion data, then data collected using assessments that rely heavily on retrospective recall should be interpreted cautiously. Finally, if recall bias is present, clinicians attempting to understand functional relations of symptoms are faced with three options: relying on biased data for clinical inferences, conducting time-consuming analogue assessments with questionable external validity, or requiring patients to self-monitor their behavior during treatment.

Hypotheses

*Hypothesis 1: OCD patients' retrospective recall of OCD symptoms will differ significantly from EMA criterion data.*

Current literature points to widespread memory biases in varied populations across numerous behaviors. Although not universal, most recent and all EMA studies that compared retrospective assessment to self-monitoring criterion data found evidence of inaccurate retrospective recall. As can be seen in Table 1, effects were observed across studies with various sample sizes, target behaviors, and lengths of recall periods. Depending upon the study, participants' retrospective recall was either overestimated or underestimated dimensions of their own behavior when compared to criterion data. With regard to OCD, experimental studies suggest that OCD patients may also have difficulty recalling dimensions of their own behavior accurately. For instance, OCD patients provided overgeneral memories when instructed to generate specific, personally relevant memories (Wilhelm et al., 1997) and performed poorly in experimental tasks involving non-verbal stimuli (e.g., stimuli from the Wechsler Memory Scales) relative to control groups (Muller & Roberts, 2005). Therefore, we hypothesized that both the estimated mean frequencies and durations of behavior provided by participants during a retrospective recall task would be statistically significantly different from EMA criterion data.

*Hypothesis 2: Ecological Momentary Assessment of OCD symptoms will not result in statistically significant reactivity.*

EMA studies have not detected reactive changes in behavior as a function of the methodology (e.g., Hufford et al., 2002; Stein & Corte, 2003). In contrast, reactivity effects have been reported in paper-and-pencil self-monitoring studies (Korotitsch & Nelson-Gray,

1999). This study examined reactivity using repeated measurements (Times 1, 2, and 3) of OCD (Y‑BOCS and OCI-R) and depression (BDI-II) symptoms. Since this study used EMA to monitor OCD but not depression symptoms, we hypothesized that reactivity effects (should they occur) would be evident in changes in participants' reports of OCD symptoms subsequent to self-monitoring with a handheld computer. The same effect would not be seen with participants' depression symptoms. We hypothesized, however, that the EMA self-monitoring would not lead to changes in participants' reports of OCD symptoms.

*Hypothesis 3: Participants' estimations of symptom covariation magnitude will be statistically significantly different from covariation statistics derived from EMA data.*

The third hypothesis examined OCD patients' ability to judge functional relations. A similar task posed to trained graduate students showed poor estimation of symptom covariation in a laboratory task (O'Brien, 1995). This study asked participants to estimate the correlation between select OCD symptoms and supplemental, non‑OCD variables. We expected OCD patients' mean estimated correlations to be statistically significantly different from correlations calculated from EMA criterion data.

Method

*Overview*

The study included 35 individuals diagnosed with obsessive-compulsive disorder who sought treatment at an anxiety disorders clinic.  Participants were first contacted by phone and asked a series of demographic questions (Time 1) and completed standardized assessments.  One week later, they met with research team members, completed additional demographic questions, completed standardized assessments, and learned how to use the handheld computers (Time 2).  At the end of the following week, they returned the handheld computers, completed standardized assessments, estimated frequencies and durations of their behavior over the previous week, and completed an online measurement tool designed to assess their ability to estimate functional relations (Time 3). Greater detail regarding experimental procedure during each point of contact with participants is provided below.

*Participants*

*Recruitment Strategies.*  Participants were recruited from the University of Michigan Health System's Anxiety Disorders Clinic in compliance with HIPPA regulations. Three recruitment strategies were utilized. First, the research project was announced to all mental health professionals in the Department of Psychiatry. Therapists were requested to describe the study to their OCD patients and obtain permission for a researcher to contact patients with further details. Second, OCD patients who expressed a willingness to participate in research at the time of their therapeutic intake procedure were identified. Finally, flyers announcing the study were posted in the University Hospital and on the electronic blackboard for clinical research on the website of the University of Michigan Health System.

*Sample.* A total of 112 patients were identified as potential participants. Of these patients, researchers were able to contact 75. Fifteen were eliminated because they did not meet diagnostic screening requirements, 16 because they changed their mind about participating in research, and one because of age restrictions. The remaining 43 patients were enrolled in the study. Seven participants dropped out of the study after initial enrollment and one participant did not complete the study due to failure of a handheld computer. Of the seven participants who dropped out of the study, two participants indicated that they did not wish to continue because they did not have enough time; the other five could not be contacted to determine the reason for their attrition.

The final sample consisted of 35 treatment-seeking patients with OCD, 48.57% ($n =$ 17) of whom were female and 51.43% ($n = 18$) male. The sample was mostly Caucasian ($n =$ 29, 82.86%), with Asian ($n = 4$, 11.43 %), African-American ($n = 1$, 2.86%), and Hispanic ($n = 1$, 2.86%) individuals also represented. The mean age of the sample was 36.31 years ($SD =$ 12.21 years, *Range* = 20 to 62) with $M = 15.88$ years of education ($SD = 2.46$). Participants were provided a $25 gift card to a local retailer in exchange for their participation.

*Measures*

*Demographic Inventory.* Appendix A includes a demographic questionnaire developed for this study. The demographic questionnaire assessed standard demographics and information relevant to inclusion and exclusion criteria. In order to reduce participant fatigue, only items necessary for making inclusion and exclusion decisions were administered at Time 1. The remainder of the items were administered at Time 2.

*Obsessive-Compulsive Inventory – Revised (OCI-R; Foa et al., 2002).* Appendix A also includes the OCI-R. The OCI-R is the short form of the Obsessive-Compulsive

Inventory (OCI) (Foa et al., 2002; Foa, Kozak, Salkovskis, Coles, & Amir, 1998). The OCI-R contains six subscales, each measuring a facet of OCD (i.e., washing, obsessing, hoarding, ordering, checking, and neutralizing). Each subscale consists of three items, for a total of 18 items. Using a 5-point Likert-scale, each question queries how much the experience has distressed or bothered you during the past month. Internal consistency for the OCI-R subscales ranged from .82 to .90 with 215 OCD patients. Two-week test-retest reliability for the subscales ranged from .74 to .91 (Foa et al., 2002). OCI-R scores were significantly correlated with other measures of OCD, including the Y-BOCS ($r = .53$), the Global Obsessive-Compulsive Scale ($r = .66$), and the Maudsley Obsessive-Compulsive Inventory ($r = .85$). OCI-R scores are also significantly correlated with measures of depression, including the Hamilton Rating Scale for Depression ($r = .58$) and the Beck Depression Inventory ($r = .70$) (Foa et al., 2002). Although there is empirical support for convergent validity for the OCI-R, discriminant validity relative to depression has yet to be demonstrated in an anxiety population. Using a total cutoff score of 18, the OCI-R discriminated OCD patients from anxious controls with 74.0% sensitivity and 75.2% specificity. To properly contextualize questions for this study, the reference time-frame for questions was changed from the past month to the past week. Further, the OCI-R was adapted to a PDA format with time intervals shortened to the previous four hours (see Appendix B).

*Yale-Brown Obsessive-Compulsive Scale (Y-BOCS; Goodman et al., 1989).* Appendix A includes the Y-BOCS, a clinician-administered semistructured interview designed to measure severity of obsessions and compulsions over the past week. The Y-BOCS consists of 10 items, five each for obsessions and compulsions. The interview is designed to follow the 64-item clinician-administered Y-BOCS checklist, which includes a

list of current symptoms (over the previous 30 days). This list is then used to identify the most prominent symptoms. Administration of the Y-BOCS takes approximately 30 minutes, with subsequent administrations taking less time. One week test-retest reliability for the Y-BOCS was 0.90 (Kim, Dysken, & Kuskowski, 1990) and interrater reliability was $r = 0.98$ (Goodman, Price, Rasmussen, Mazure, Fleischmann et al., 1989). Internal consistency alpha coefficients have ranged from 0.88 (Goodman, Price, Rasmussen, Mazure, Fleischmann et al., 1989) to 0.69 (Woody, Steketee, & Chambless, 1995). The Y-BOCS has been shown to correlate significantly with other measures of OCD. However, as with the OCI-R, the Y-BOCS correlates strongly with measures of general anxiety and depression (Goodman, Price, Rasmussen, Mazure, Delgato et al., 1989).

*Beck Depression Inventory-II (BDI-II; Beck et al., 1996).* [The BDI-II is not included in an appendix because Harcourt Assessment, Inc., denied permission to publish in full or in part.] The BDI-II is a 21-item self-report questionnaire that measures the presence and severity of depressive symptoms and suicidal ideation (Beck, Steer, & Brown, 1996). Administration takes approximately 5 minutes. One week test-retest reliability for the BDI-II is 0.93 and internal consistency was 0.92 among outpatients (Beck et al.). The BDI-II correlated highly with other measures of depression (Beck et al.) but, like most measures of mood, had difficulty discriminating between depression and anxiety (Dozois & Dobson, 2002). Use of the BDI-II served two purposes. First, it provided an index of depression and was used as a covariate to assess whether retrospective recall accuracy is a function of depressive symptomatology (Wilhelm et al., 1997). Second, parallel assessment of depressive and OCD symptomatology was used to estimate reactivity effects to the EMA data collection method by providing a non-EMA data collection control.

*Anxiety Disorders Interview Schedule for DSM-IV (Adult Version) (ADIS-IV).* The

ADIS-IV is a semi-structured assessment interview for differential diagnosis of DSM-IV

anxiety disorders and other Axis I disorders. The ADIS-IV possesses good inter-rater

reliability, with kappa coefficients between .60 and .86 across disorders, except for

Dysthymic Disorder (kappa = .22) (Brown, DiNardo, Lehman, & Campbell, 2001). The

ADIS-IV was used to increase sample homogeneity by ensuring that participants were

accurately diagnosed with OCD during the screening procedure and to assess for comorbid

conditions. Subsections not relevant to the study were excluded, a procedure believed not to

compromise the psychometric integrity of the interview as a whole (Brown, DiNardo, &

Barlow, 1994). Therefore, the following subsections were excluded to reduce participant

fatigue: Dysthymic Disorder, Hypochondriasis, Somatization Disorder, Mixed Anxiety-

Depression, Medical History, and the Hamilton Scales for Depression and Anxiety.

*Assessors*

Assessors were advanced doctoral clinical psychology students and a clinical social

worker (MSW). Training for this project entailed several steps. First, all assessors watched a

training video of the clinician-administered interviews (Y-BOCS and ADIS-IV). Each video

contained several vignettes. Assessors were required to agree with at least 80% of the

experts' (Initials J. H. and E. K.) diagnostic decisions and total scores on the Y-BOCS and

ADIS-IV. All discrepancies from the criterion were discussed, and the rationale for the

correct rating was provided by the principal investigator. Each assessor observed one Y-

BOCS and ADIS-IV interview before administering supervised interviews. During these

observations, ratings between the principal investigator and the trainees were compared and

disagreements were discussed. If rating discrepancies between the principal investigator and

the trainees exceeded 1 point on either the Y-BOCS total score or ADIS-IV diagnostic

severity scale, trainees were required to review the training videos.

*Power Analysis*

Estimates of predicted effect sizes regarding the magnitude of discrepancies between

retrospective memory estimates and EMA data were based upon previous studies of

reasonable similarity. Table 1 reports effect sizes of relevant studies. Overall, the mean

Cohen's *d* for EMA studies similar to the current study was 0.60 (*SD* = .43, Range = -0.16 to

0.97). Because previous effect sizes have been variable, a projected Cohen's *d* of 0.50

appeared justified. For dependent samples *t*-tests with an anticipated effect size of 0.50 and

alpha set to .05, 35 participants were required for statistical power of .82 (Borenstein,

Rothstein, & Cohen, 1997). In comparison, a study of similar design and purpose that

examined retrospective recall accuracy of eating-disordered behaviors reported a sample size

of sixteen participants (Stein & Corte, 2003).

*Inclusion and Exclusion Criteria*

Inclusion criteria for the study were (a) age 18 to 65, (b) OCD diagnosis according to

the ADIS-IV, and (c) a score of 16 or higher on the Yale-Brown Obsessive-Compulsive

Scale (Y-BOCS) or greater than 18 on the Obsessive-Compulsive Inventory – Revised (OCI-

R). Exclusion criteria were (a) age under 18 or over 65, (b) co-morbidity with a disorder

known to be associated with impaired memory (i.e., current psychotic disorder, current

substance dependence, mental retardation, dementia, bi-polar disorder), (c) any physical

disabilities that prohibited participation (i.e., inability to view the computer screen, hear the

beep of a hand-held computer, or enter data into the computer), (d) inability to understand

English, (e) suicidality, and (f) subclinical anxiety symptoms as determined by a Y-BOCS

score below 16 or an OCI-R score below 18.

*Procedure*

The study was conducted over a two-week period with three observation points. Time

1 observations occurred on the first day of participation. Time 2 observations occurred seven

days later, on the eighth day of the study. Time 3 observations occurred seven days after that,

on the 15th day of the study. Each point of contact with participants (i.e., Time 1, Time 2, and

Time 3) is described in more detail below. Figure 1 shows the design visually.

*Screening Assessment (Time 1).* Prospective participants were screened by telephone.

Patients were informed of their rights as a research participant and oral informed consent was

solicited prior to data collection. During the telephone screening, participants answered

demographic questions and were orally administered the OCI-R, Y-BOCS, and BDI-II.

Questions were modified when necessary to query for symptoms over the last week.

Participants meeting inclusion criteria were scheduled for EMA training at Time 2. Finally,

participants were instructed to informally monitor their symptoms during the ensuing week

with the statement: "Please pay attention to your symptoms over the next week."

*Time 2 (EMA Training).* One week later participants met with a researcher and signed

the informed consent. Participants completed the OCI-R, Y-BOCS, and BDI-II a second

time. Instructions for the OCI-R were modified to query for symptoms occurring in the

previous week. In order to confirm participants' OCD diagnosis and identify comorbid

conditions, participants also completed the ADIS-IV. An estimate of the accuracy of each

participant's one-week memory was established by having participants recall their answers

from the OCI-R at Time 1. Thus, difference scores between Time 1 OCI-R scores and Time 2

recall of those same scores served as an index of each participant's ability to recall symptom report information over a seven-day interval.

Participants then received a Palm Zire 21 hand-held computer and were instructed in its use. The handheld computer had a monochrome, 5cm by 5cm screen. Data were recorded by touching the screen using a stylus. Instruction included a demonstration of (a) how to operate the hand-held computer, (b) how to initiate data recording, (c) how to correctly record data, and (d) the beep that would signal them to record. Participants had the opportunity to practice recording data and to ask questions, and were told the PDA would time stamp their responses. Participants were also informed that every time they completed an assessment within 15 minutes of being signaled, their name would be entered in a lottery giveaway of a PDA at the end of the study. Participants were instructed to record any technical difficulties or malfunctions in a notepad included with the PDA. A final appointment was scheduled for seven days later. Participants were contacted on the evening of the first day of EMA self-monitoring to address any questions regarding the use of the PDA. Participants were told to call a researcher at any time if they had questions. The research number was affixed inside the notepad attached to the hand-held computer. In addition, participants were contacted on the third day to assess for any difficulties and again on the sixth day to remind them of their appointment the following day.

*EMA Self-Monitoring.* During the week between Time 2 and Time 3 (days 8 to 14 of the study) participants completed questionnaires on the PDA contingent upon an audible signal tone every four hours (i.e., 10 a.m, 2 p.m, 6 p.m., and 10 p.m.) Each item read, "In the last four hours…" and was worded in the past tense. Items delivered via EMA were adapted from the OCI-R. Selected items were chosen because of their strong OCI-R factor loadings

and because they represented six important facets of OCD (i.e., washing, obsessing, hoarding, ordering, checking, and neutralizing) (Foa et al., 2002). Each item consisted of a screening question followed by two follow-up questions if the screening item was endorsed. First, the participant was asked whether he or she had engaged in a target behavior during the previous four hours.  If the participant did not endorse the item, the program algorithm branched to the next OCD item.  However, item endorsement caused the program to branch to two follow-up questions that assessed the frequency and duration of the symptom.  After all 18 OCD items were assessed, the PDA administered seven supplemental items that assessed length of sleep the previous night, present physical location, whether the participant was alone, mood, stress level, anxiety, loneliness, and whether he or she experienced distress following subjectively defined interpersonal conflict during the previous four hours. These additional items were used for analyses involving the ability of participants to detect functional relations and are hereafter referred to as supplemental items.

Sampling procedures available to EMA include (a) event contingent, or initiation of recording based on the occurrence of a behavior; (b) fixed interval, or initiation of recording following auditory prompts from the computer that occur on a predetermined regular time interval (e.g., every 2 hours); and (c) random interval prompts, or initiation of recording following auditory prompts from the computer that occur on a random interval schedule (Shiffman, 2000; Stone & Shiffman, 1994). Interval sampling was used because event-contingent sampling is best suited for discrete episodes of low frequency behaviors. The diffuse nature of some obsessions would have rendered it impossible for participants to determine when one obsessive episode ended and another began. Random interval sampling was not used because retrospective recall intervals would not have been constant across

participants, thus introducing a potential experimental confound. Combining different types of sampling procedures (e.g., concurrent fixed and event-contingent sampling) was conceivable; however, previous attempts with bulimia nervosa patients failed to generate meaningful quantities of data from event-contingent entries (Wunderlich, December 9, 2004, personal communication). Because a fixed interval sampling method provided both statistical and practical advantages, the PDA prompted participants four times per day (10 a.m., 2 p.m., 6 p.m., and 10 p.m.). Hourly prompting was rejected based on participant fatigue reported by Herman and Koran (1998).

*Time 3 (Retrospective Recall and Estimation of Functional Relations).* On the last day of the study, participants returned the PDA and completed the OCI-R, Y-BOCS, and BDI-II. Using the prior week as the context, participants estimated the total frequency and average duration of each OCD symptom queried by the PDA over the preceding week. For example, participants were asked, (a) "In the last week (when you used the handheld computer), how many times did you…" and (b) "On average, how long did you spend doing … each day?" Retrospective estimates of daily behavioral frequencies and durations were then compared to EMA aggregated data of mean daily behavioral frequencies and durations. The difference between a participant's retrospective estimate and the EMA collected data was the index of recall accuracy.

After estimating the frequency and duration of all EMA-OCD items, participants then completed a PowerPoint tutorial that explained the concept of correlation (see Appendix C). The purpose of the tutorial was to teach participants the meaning of co-variation so that they could estimate functional relations between selected OCD symptoms and other variables (e.g., amount of sleep, mood, etc.). Embedded in the tutorial were concept quizzes on the

topics of positive correlation, negative correlation, zero-order correlation, strong correlation, and weak correlation. Participants were required to successfully answer these questions before continuing. The participants then estimated correlations between select OCD symptoms and supplemental EMA items encountered during the past week. Appendix C includes the correlation tutorial and the scales used to assess participants' accuracy in estimating covariance of their symptoms with supplemental items. Participants estimated correlations with the aid of a color-coded visual analogue scale ranging from -100 (perfect negative correlation) to +100 (perfect positive correlation). We chose to express correlations as a range from -100 to +100 in order to simplify the task for participants who might not understand the meaning of decimal fractions.

The algorithm for selecting EMA-OCD items used in the correlation estimation task was as follows. Given that the EMA-OCD was adapted from the OCI-R, participants' OCI-R scores assessed at Time 2 were used to select items for the correlation estimation task. In order to broaden the range of behaviors in the estimation task, only one item from each subscale was selected for each participant. Using data from Time 2, the three highest OCI-R subscales were selected. Within each of these subscales, the one item endorsed at the highest level was selected for the correlation estimation task. Occasionally, participants endorsed an OCI-R item during the Time 2 assessment but did not endorse the corresponding EMA-OCD at all during the following week. For such occurrences, up to three additional items (one for each occurrence) endorsed during the Time 3 retrospective recall task were selected in an effort to provide participants non-zero items for the correlation estimation task. Further, assessors were given the latitude to assess additional high-frequency or high-duration items revealed during the Time 3 retrospective recall task, if they were not captured by the OCI-R

at Time 2. Thus, participants estimated correlations between the supplemental EMA items for between 3 and 6 OCD behaviors.

Participants were then asked how confident they were in their memory, how accurate they were in responding to PDA questions, and to estimate the accuracy of their covariation judgments. Finally, they were asked about their awareness of the study's hypotheses. Participants were then debriefed and paid $25 in the form of a gift certificate to a local retailer. In addition, participants who had consistently responded to PDA prompts in a timely fashion had their names entered into a drawing for a new hand-held computer.

Results

*Participants' Symptom Severity*

During the screening interview at Time 1, participants completed the OCI-R, Y-BOCS, and BDI-II. Means scores for all participants included in the study ($N = 43$) were as follows: OCI-R total score = 26.14 ($SD = 9.67$), Y-BOCS severity score = 22.44 ($SD = 3.98$), and BDI-II total score = 22.67 ($SD = 12.70$). At Time 2, all participants completed the ADIS-IV to verify their Axis I diagnoses. As measured by the ADIS-IV, the mean OCD severity level was 5.44 ($SD = 0.81$). The ADIS-IV uses a 0 to 8 scale with 4 considered the cutoff for diagnostic presence. In addition to a principle or co-principle diagnosis of OCD, the mean number of comorbid diagnoses in the sample was 1.31 ($SD = 1.13$). Social Phobia was the most frequent comorbid diagnosis ($n = 10$), followed by Major Depression ($n = 9$), Panic/Agoraphobia ($n = 8$), Generalized Anxiety Disorder ($n = 4$), Specific Phobia ($n = 2$), Posttraumatic Stress Disorder ($n = 1$), Dysthymia ($n = 1$), Body Dimorphic Disorder ($n = 1$), and Trichotillimania ($n = 1$).

*Data Screening*

*Missing data.* The effect of missing data in an experiment of this type can compromise inferences regarding the accuracy of retrospective recall in participants, given that the EMA data form the criterion by which accuracy is assessed. Missing criterion data can have the effect of degrading the quality of the criterion so that it is not a veridical reflection of actual behavior. Thus, careful consideration of procedures for handling missing data is required. Therefore, Hypothesis 1 was examined first without employing statistical corrections to account for missing data. It was decided that missing data corrections would follow the first set of analyses if statistically significant differences were detected between

EMA values and retrospective recall. In this way, the robustness of the findings would be tested to determine if missing data accounted for any of the findings.

If a participant in this study completed the entire protocol, he or she would have completed the EMA questionnaires 28 times over seven days. Unfortunately, the demands of repeated measurement increase the likelihood that at least one (and perhaps more) EMA observations will be missed. Generally, missing data can be categorized into three theoretically derived but untestable statistical relationships: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (King & King, 2001; Schafer & Graham, 2002). Missing data are considered ignorable if they are MCAR or MAR. In this data set, the theoretical argument for the existence of MCAR or MAR is likely. In other words, the probability that an EMA observation is missing is unrelated to the value of the EMA-OCD variables or to the supplemental EMA variables.

Furthermore, in a repeated measures design, data can be missing at the unit level (e.g., a participant does not answer any questions), item level (e.g., a participant answers some questions during a given observation, but not others), and wave level (e.g., a participant fails to answer one or more observations, but continues in the experiment to answer subsequent observations). In this study, item level omissions did not occur because the EMA program did not allow participants to continue without answering all questions. Instead, missing data occurred at the wave level and consisted of scheduled observations that participants' failed to record (e.g.., when a study completer missed a 10 a.m. assessment, but completed a subsequent assessment). Although procedures to correct for wave level missing data are less established than item level data (Schafer and Graham, 2002), a commonly accepted practice is to estimate missing data based on qualitatively similar waves (e.g.,

within participant) to account for missing waves. Random-effects regression accommodates for missing data in this way and is recommended for studies with multiple measurements nested within participants (King & King, 2001).

*Evidence of data entry errors or outliers.* Prior to analysis, these data were examined through various SPSS programs for accuracy and outliers. First, impossible values (e.g., recording 700 minutes as the duration of a behavior in a 4 hour period) were identified. Given the nature of recording on a handheld computer, inadvertent recording errors were also possible. Therefore, inconsistencies within participants across the week were identified. In this study, an outlier was defined as a data point that was at least ten times greater than the mean value for the same item for any given participant over the course of the week. Such an outlier could have occurred due to inadvertent replication of a number on the keypad. Three participants recorded impossible values, one of whom utilized an idiosyncratic recording strategy (i.e., recorded 9999999 for frequency and duration consistently across the week and in her recall). The other two participants appeared to inadvertently record impossible values. When previously recorded values clearly pointed to recording errors (i.e., 120 minutes recorded consistently across the week followed by 1200 minutes), the impossible value was changed to be consistent with the rest of that participant's data. This occurred 4 times and was confirmed by written comments provided by two participants in the notebook provided with the handheld computer. The remaining four outliers were changed to zero to prevent the artificial inflation of EMA values, thereby guarding against data cleaning procedures that favor the experimental hypothesis.

*Response Rate*

Overall, participants responded to 872 of a possible 969 EMA observations for an overall response rate of 90.0%. Of these, 702 (80.5%) occurred within 15 minutes of the audible signal tone (either before or after), 752 (86.2%) within 30 minutes, and 798 (91.5%) within 60 minutes. The mean latency between signal and response was 21.14 minutes ($SD$ = 54.89). The response rate for this study is compatible to reported response rates of previous studies with a similar design (i.e., 94%, Stone et al., 2004).

*Hypothesis 1: OCD patients' retrospective recall of OCD symptoms will differ significantly from EMA criterion data*

At Time 3, participants were asked to recall the total frequency of behaviors measured by the EMA-OCD items over the previous week. Two analytic strategies were used to measure participants' retrospective recall accuracy relative to the EMA criterion data. The first data analysis strategy utilized paired-samples *t*-tests to test for statistically significant differences between retrospective recall estimates at Time 3 and aggregated EMA data. The second strategy, random-effects regression, supplemented the *t*-tests and took into account the missing wave-level data. The random-effects regression analysis was conducted in order to take full advantage of the repeated observations within individuals. Implications of these different approaches are addressed in the discussion section.

*Paired-samples t-tests*

A series of paired-samples *t*-tests were conducted to evaluate participants' ability to accurately recall the frequency and duration for each of the 18 EMA-OCD recorded behaviors the previous seven days. Given the idiosyncratic symptom profile of OCD, only a subset of the 18 EMA-OCD items were relevant for each participant. Participants who

neither endorsed an item during EMA monitoring nor during the Time 3 recall were excluded

prior to conducting the analyses for that item. Therefore the *n* for each item was contingent

upon the number of participants who endorsed the item during the previous week. As noted

above, one participant was excluded due to failure to conform to the study's protocol.

Tables 2 and 3 display the means, standard deviations, and statistical comparisons for

each of the eighteen frequency and duration EMA-OCD items. Using a Bonferonni family-

wise correction, the traditional .05 level of significance was divided by the number of

analyses run (one for each of the 18 EMA-OCD items), thereby setting alpha for each

comparison at .002 and the experiment-wise alpha at .05. As shown in Tables 2 and 3, none

of the comparisons resulted in a statistically significant difference between the participant's

estimate and the EMA-OCD criterion. However, Cohen's paired samples *d* suggested that the

difference between the EMA criterion values and retrospective recall resulted in medium to

large effect sizes for some behaviors. With respect to frequency questions, all items with an

effect size $\leq .3$ were underestimates of the behavior when retrospectively recalled relative to

the EMA criterion values. With respect to duration questions, all items with an effect size $\geq$

.3 were overestimates of the behavior when retrospectively recalled relative to the EMA

criterion values. Taken together, these results suggest that when asked to recall

ideographically relevant OCD behaviors, participants were often relatively accurate when

compared to an EMA criterion value. When inaccurate, however, it appears as if participants

underestimated with respect to frequency, but overestimated with respect to duration.

Combined, this suggests that participants exaggerate their duration per behavior recall for

some behaviors. However, it is important to note that Cohen's *d* is mathematically related to

the paired samples *t*-test[2]. Thus, given the limitations of the *t*-test noted below, conclusions based on effect sizes should be interpreted cautiously.

Given that none of the comparisons yielded significant differences, analyses were not repeated with missing data because replacement of missing EMA waves would only reduce differences between retrospective recall and the criterion values. However, as can be seen in Tables 2 and 3, the non-significant findings occurred despite large mean differences. Removal of participants from item-level analyses who did not endorse that behavior resulted in reduced power and contributed to the non-significant findings. It is thus possible that the nature of these data may be inappropriate for the paired-samples *t*-test. One reason is that aggregating the repeated EMA observations reduced statistical power and increased the risk of Type II error. Further, the denominator in the paired-samples *t*-test captured all sources of error variance in a single term. This approach did not capitalize on the repeated measures nature of the data, which allows error to be partialed into separate terms. The failure to do so also leads to the increased risk of Type II error. In comparison, random-effects regression utilizes the full nature of repeated observations nested within individuals. With such data, random-effects regression reduces Type II error and provides more accurate inferential conclusions (Hedeker, 2004; Rowland & Thornton, 2003).

*Random-effects regression*

Hypothesis 1 was also tested using random effects regression with a feasible generalized least square (FGLS) estimator. FGLS accounts for autocorrelation, thus yielding unbiased and efficient estimates of recall accuracy. Dummy variables were coded to allow for statistical testing of the effect of the variable in question (e.g., retrospective recall vs. EMA mean). The equation estimated for the measurement $\gamma$ of subject $i$ on occasion $t$ is:

---

[2] Cohen's *d* for paired samples = paired-samples $t/\sqrt{N}$

$$\gamma_{it} \quad = \quad \alpha \quad + \quad \beta_1 X_{it} \quad + \quad \beta_2 S_i \quad + \epsilon_{it}$$

where,

| | | |
|---|---|---|
| $i$ | = | 1, 2, …, $N$ subjects; |
| $t$ | = | 0, 1, …, $T$ timepoints; |
| $\alpha$ | | is the constant (overall population intercept); |
| $\beta_{1, 2, etc.}$ | | is the coefficient for that variable; |
| $X_{it}$ | | is the dummy variable to be tested (e.g., EMA = 0, retrospective recall = 1) |
| $S_i$ | | is the value for subject variables (e.g., education, memory, etc.) for subject $i$; and |
| $\epsilon_{it}$ | | is the independent error for subject $i$ at time $t$. |

Furthermore, the error term $\epsilon_{it}$ can be broken down into $u_i$ (unexplained variance that varies across participants but not over time) and $e_{it}$ (unexplained variance that varies across participants and time). This procedure capitalizes on the richness of panel data, accounts for unobserved variables that vary across individuals, and thereby decreases Type II error. Finally, this procedure is one of the most efficient and accurate means of accounting for missing data (King & King, 2001). For these analyses, participants' EMA-OCD data were summed per day for each participant with complete data. Linear interpolation was used to account for missing data within individuals for participants who missed one or more EMA observations.

*Rationale for Predictor Variables.* The first set of random-effects regression analyses examined participants' retrospective recall accuracy with respect to behavioral frequency after accounting for several theoretically relevant variables. The flexibility to include additional explanatory variables in the random-effects regression reduced unexplained variation of the OCD symptoms, thereby yielding more precise estimates of the coefficient of recall and decreasing the probability of a Type II error (Rowland & Thornton, 2003). The following eight variables were included in the model because of the potential influence on OCD symptoms: (a) the number of years participants reported OCD symptoms, (b) OCD

severity as measured by the ADIS-IV at Time 2, (c) a proxy test for one-week memory, (d) self-reported accuracy of their retrospective recall following the recall task, (e) prior knowledge of Hypothesis 1, (f) depression severity as measured by the BDI-II, (g) sex, and (h) age. The first variable was included in the model based on the assumption that symptom stability would be positively correlated with the amount of time the participant had experienced the symptom. OCD severity was included on the assumption that OCD severity may influence OCD symptom variance. The memory proxy, calculated as the difference between OCI-R values assessed at Time 1 and recall of those answers at Time 2, was included based on the assumption that participants with worse scores may have had more difficulty recording their symptoms via EMA. Both participants' self-reported accuracy of the retrospective recall task and knowledge of the experimental hypothesis are potentially related to OCD symptoms and were therefore included in the model. Depression severity was included in the model based on the results of a previous study in which apparent memory deficits in OCD disappeared once the effect of depression was accounted for (Wilhem, McNally, Baer, and Florin (1997). The variable sex and age were included as standard demographic variables. Given that the primary goal of this analysis is to examine the accuracy of participants' retrospective recall accuracy, these variables will not be interpreted individually. Instead, they serve as the context for providing precise estimates of the coefficient of recall.

*Behavioral Frequency.* Coefficients for the dummy variable recall (EMA vs. retrospective recall of behavioral frequency) are presented in Table 4 along with inferential statistics for each EMA-OCD item's model. The coefficient of recall indicates the estimated difference in the average daily frequency between recalled and EMA data. In these models,

the coefficient of recall represents the average change after accounting for the eight predictor variables described above, partialing the error variance, and autocorrelation. The error variance is partialed into two components: (a) unexplained variance that varies across subjects but not over time ($u_i$) and (b) unexplained variance that varies across subjects and time ($e_{it}$). As can be seen in Table 4, most of the coefficients are negative values, indicating that compared to EMA calculated average daily frequency, participants' retrospectively recalled average daily frequencies were underestimated. Across the 18 EMA-OCD items, the average difference between retrospectively recalled daily frequencies and EMA daily frequencies was generally less than 10, suggesting relatively accurate estimation. However, two items were statistically significant: item 2 (i.e., "I checked something I didn't need to") and item 10 (i.e., "I repeated certain numbers"). Of these, only the coefficient of recall for item 10 suggested meaningful differences. The coefficient of recall indicates that after accounting for the predictor variables, participants retrospectively recalled 139.3 more incidences per day of "repeating certain numbers" than what was recorded on the handheld computer.

*Behavioral Duration.* The second set of random-effect regression analyses examined participants' retrospective recall accuracy with respect to average daily behavioral duration of the OCD behaviors after accounting for the same variables identified above. Coefficients for the dummy variable recall (EMA vs. retrospective recall of behavioral duration) are presented in Table 5 along with inferential statistics for each EMA-OCD item's regression model. As can be seen in Table 5, only EMA-OCD items 4 and 10 resulted in a significant coefficient of recall. With the exception of these two items, differences between retrospectively recalled average daily duration and EMA values did not exceed 15 minutes.

This suggests that after accounting for the predictor variables, participants are relatively accurate when retrospectively recalling average daily behavioral durations. The two EMA-OCD items with significant coefficients of recall indicated that participants' retrospective recall of average daily durations was overestimated by 36 minutes for repeatedly counting objects (item 4), $z\,(125) = 2.88$, $p < .004$, and 28 minutes for repeating certain numbers (item 10) $z\,(100) = 2.44$, $p < .05$.

The implication of these results is that, on average, OCD patients appear to accurately recall both the frequency and duration of most OCD symptoms. Although explanation of symptom variation was not a primary goal of this analysis, Tables 4 and 5 provide information about the overall predictability of OCD symptoms, given the assumptions of this model. The models' overall $R^2$ for behavioral frequency suggests that at best (EMA-OCD item 16), approximately 50% of the symptom variance is explained by the variables in these models. However, most of the models $R^2$ were below .37, suggesting that for most items, approximately 63% of the variation in OCD symptoms remain unexplained. Similarly, $R^2$ values for the models of behavioral duration were all below .37.

Further, random effects regression yields information regarding possible sources of unexplained variance. More specifically, Rho indicates the percentage of unexplained variation in symptoms due to unobserved factors that differ across subjects (i.e., subject specific factors), but are constant over time. By way of extension, variance not accounted for by subject specific factors is due to unobserved factors that differ across subjects and time. For both behavioral frequency and duration models, Rho statistics were relatively high for items measuring obsessions (i.e., EMA-OCD items 6, 12, and 18). This suggests that relative to other items, unexplained variance in items measuring obsessions are likely accounted for

by participant specific factors that do not vary across time. These results suggest avenues of investigation regarding the prediction of symptoms for future studies.

*Hypothesis 2: Ecological Momentary Assessment of OCD symptoms will not result in statistically significant reactivity*

In order to assess for the presence of reactivity, one-way within-subjects ANOVAs were conducted across the three time points using total scores from the OCI-R, Y-BOCS, and BDI-II as dependent variables. The mean and standard deviations for these measures are presented in Table 6. With respect to the OCI-R total score, results indicated a significant time effect, Wilks's $\Lambda = .65$, $F(2,33) = 8.78$, $p<.001$, multivariate $\eta^2 = .35$. Follow-up polynomial contrasts indicated a significant linear effect with the mean OCI-R total score decreasing significantly from Time 1 to Time 2 but not from Time 2 to Time 3.

Given the significant time effect observed in the OCI-R total score, one-way within-subjects ANOVAs were repeated for each of the OCI-R subscales. Significant time effects were detected for four of the six subscales: Washing, Wilks's $\Lambda = .77$, $F(2,33) = 4.93$, $p<.013$, multivariate $\eta^2 = .23$; Obsessing, Wilks's $\Lambda = .69$, $F(2,33) = 7.28$, $p<.002$, multivariate $\eta^2 = .31$; Hoarding, Wilks's $\Lambda = .63$, $F(2,33) = 9.89$, $p<.001$, multivariate $\eta^2 = .36$; and Checking, Wilks's $\Lambda = .79$, $F(2,33) = 4.45$, $p<.019$, multivariate $\eta^2 = .21$. Significant time effects were not detected for the subscales Ordering, Wilks's $\Lambda = .97$, $F(2,33) = 1.31$, $p<.28$, multivariate $\eta^2 = .07$ or Neutralizing Wilks's $\Lambda = .92$, $F(2,33) = 1.51$, $p<.24$, multivariate $\eta^2 = .08$. Follow-up polynomial contrasts indicated significant linear effects for each of the four statistically significant subscales. Pairwise comparisons revealed statistically significant time effects between Time 1 and Time 3 for each of the subscales. Statistically significant time effects were also detected between Time 1 and Time 2 for the Obsessing,

Hoarding, and Checking subscales, but not for the Washing subscale. In contrast, time effects were not detected between Time 2 and Time 3 for any of the subscales. In conjunction with an examination of the means shown in Table 6, these results suggest that the time effects noted in the OCI-R total score and subscales are due to changes between Time 1 and Time 2 but not Time 2 and Time 3.

Reactivity of the Y-BOCS was also examined using a one-way within-subjects ANOVA. In contrast to the OCI-R, significant time effects were not detected for Y-BOCS total score, Wilks's $\Lambda = .85$, $F(2,33) = 2.93$, $p<.067$, multivariate $\eta^2 = .15$. Given *a priori* hypotheses and results from the OCI-R obsession subscale, the Y-BOCS Obsession and Compulsion subscales were independently examined for reactivity. Results indicated a time effect for the Obsession subscale, Wilks's $\Lambda = .73$, $F(2,33) = 6.14$, $p<.005$, multivariate $\eta^2 = .27$, but not the Compulsion subscale Wilks's $\Lambda = .98$, $F(2,33) = .35$, $p<.705$, multivariate $\eta^2 = .02$. With respect to the Y-BOCS Obsession subscale, follow-up polynomial contrasts indicated a significant linear effect with the mean OCI-R total score decreasing significantly from Time 1 to Time 3 and from Time 2 to Time 3, but not from Time 1 to Time 2.

Finally, reactivity was examined in the BDI-II using a one-way within-subjects ANOVA. Results revealed a significant time effect for the BDI-II total score, Wilks's $\Lambda = .61$, $F(2,33) = 10.50$, $p<.001$, multivariate $\eta^2 = .39$. Follow-up polynomial contrasts indicated a significant linear effect with a significant decrease from Time 1 to Time 2 and from Time 1 to Time 3, but not from Time 2 to Time 3. Similar to the OCI-R, most of the change in scores was observed during the first week.

Of the three possible comparisons between the three time points, only the comparison between Time 1 and Time 3 was significant for each scale, with decreasing means across the

study (see Table 7). Therefore, if a polynomial contrast between Time 1 and Time 2 was significant, the contrast between Time 1 and Time 3 is automatically significant. Given this fact and because the purpose of these analyses was to detect reactivity of informal self-monitoring of week 1 (Time 1 to Time 2) in comparison to EMA monitoring of week 2 (Time 2 to Time 3), interpretation of change per week was more important than the sum total of change across both weeks. Thus, comparing change during week 1 and week 2 suggested that most change occurred during the first week of informal self-monitoring. In fact, only the Y-BOCS Obsession Subscale showed a significant change during week 2.

*Hypothesis 3: Participants' estimations of symptom covariation magnitude will be statistically different from covariation statistics derived from EMA data.*

*Discrepancies in Estimated and EMA Correlations for Aggregated OCD Behaviors.* To test the hypothesis that participants' estimates of magnitude of symptom covariation were significantly different from those found in the EMA data, participants estimated the relationship of selected OCD items with the six supplementary variables.  For example, a participant might estimate the relationship between obsessive thinking with sleep duration, mood, stress level, anxiety, loneliness, and distress following an interpersonal fight.  After these estimates were completed, the process was reiterated for a second OCD symptom with the same supplemental variables.  Thus, each participant provided a minimum of 18 co-variation estimates (at least 3 OCD behaviors x 6 supplemental variables = 18 estimates). The procedure section above describes the algorithm used to select the EMA-OCD items.

A series of paired-samples *t-* tests were conducted to evaluate whether retrospective covariation estimates (hereafter referred to as *estimated* correlations) were statistically different from correlations calculated from the EMA data (hereafter referred to as *EMA*

correlations). Since it is unclear whether participants answered these questions based on behavioral frequencies or behavioral durations, estimated correlations were compared to EMA correlations separately for behavioral frequencies and behavioral duration. Using a Bonferonni family-wise correction, the traditional .05 level of significance was divided by the number of analyses run (one for each of the 6 supplemental items), thereby setting statistical significance at .008 for any individual comparison. Means, standard deviations, and test statistics of overall estimated and EMA correlations were collapsed across individuals, behavior, and individual rank-order of behavioral frequency and duration and are listed in Table 8.

For the following analysis, each participant contributed multiple responses, thus potentially violating the assumption of independence of observations. Violation of this assumption normally leads to an increase in Type I error. However, as will be seen below, the presence of some statistically non-significant results lend credence to the validity of these conclusions.

Results showed that participants' estimated correlations frequently departed from EMA correlations to a degree that could not be accounted for by chance or measurement error alone. In general, participants consistently overestimated the degree of relationship between OCD symptoms and variables measured by the supplemental items. Significant overestimations were found for correlations between behavioral frequency and stress $t(117) = 7.58$, $p < .001$, $d = .70$; anxiety $t(114) = 9.15$, $p < .001$, $d = .85$; and loneliness $t(101) = 3.57$, $p < .001$, $d = .35$. Significant overestimations were found for correlations between behavioral duration and stress $t(117) = 7.02$, $p < .001$, $d = .64$; anxiety $t(113) = 7.83$, $p < .001$, $d = .73$; loneliness $t(101) = 4.04$, $p < .001$, 40; and distress following subjectively defined

interpersonal fights $t(30) = 2.88$, $p < .007$, $d = .52$. It is noteworthy that the absolute

magnitude of estimated symptom covariation was always greater than the EMA symptom

covariation. In other words, participants were more likely to significantly overestimate

symptom covariation than significantly underestimate symptom covariation. Interestingly,

results indicate that the highest EMA correlation between OCD symptom fluctuation and any

supplemental item was only 0.21 (for the relationship between OCD behavioral duration and

levels of stress and anxiety) and accounted for just 4% of symptom variance. Nevertheless,

participants estimated a mean correlation of 0.51 between OCD symptoms and levels of

stress and anxiety, thereby estimating that 26% of the variation in OCD symptoms could be

explained by contemporaneous levels of stress and anxiety.

Given these findings, subsequent analyses were conducted to determine if participant-

level rank ordering of behavioral frequency and duration or specific OCD behaviors also

resulted in significant differences between participants' estimations of symptom covariation

magnitude and EMA magnitudes.

*Discrepancies in Estimated and EMA Correlations for OCD Behaviors Grouped by*

*Rank Order.* Previous research suggests that accuracy of co-variance estimation is a function

of how frequently a behavior occurs. Therefore, co-variance estimates for behaviors that

occur with high frequency may be more difficult to accurately estimate than behaviors

occurring less frequently (O'Brien, 1995). In order to examine this hypothesis, participants'

Time 3 estimates of behavioral frequency and duration were placed in rank order. Given

concerns about statistical conclusion validity for extremely small samples, statistical tests

were conducted only on behaviors ranked 1st to 4th. When participants reported identical

frequencies or durations for two or more behaviors, ties were assigned the same rank order.

Given this ranking procedure and the procedural algorithm for selecting comparisons at Time 3, only two participants estimated the covariation of six OCD behaviors, while only five participants estimated the covariation of five behaviors. Means, standard deviations, and test statistics comparing estimated and EMA symptom covariation grouped according to rank order are presented in Table 9 for behavioral frequencies and Table 10 for behavioral durations.

As can be seen in Table 9, participants' consistently overestimated covariation of the four most frequently occurring OCD symptoms with levels of anxiety (Cohen's $d = .56$ to 1.25 for anxiety) and stress (Cohen's $d = .60$ to .90). As can be seen in Table 10, the pattern was replicated with respect to behavioral duration for the supplemental items measuring levels of anxiety (Cohen's $d = .41$ to .94) and stress (Cohen's $d = .33$ to .86 for stress).

In contrast, results for both behavioral frequencies and durations pointed to participants' ability to accurately estimate their symptom covariation of the four most frequently occurring OCD symptoms with sleep duration ($d = .10 - .25$ for symptom frequency; $d = .05 - .33$ for symptom duration). Less consistent are comparisons between estimated and EMA correlations for loneliness and mood. Although statistically significant differences between estimated and EMA correlations for loneliness emerged only with respect to behavioral durations (i.e., longest duration, $t (37) = 3.55, p < .001, d = .58$ and fourth longest duration, $t (9) = 4.04, p < .001$), moderate effect sizes were observed for both behavioral frequencies ($d = .10 - .50$) and durations ($d = .22 - .58$ for stress). With respect to the supplemental item measuring level of current mood, clear patterns across the rank orders of behavioral frequencies and durations did not emerge. The largest discrepancy between estimated and EMA correlations for the supplemental variable mood with behavioral

frequencies was noted for the least frequent behavior $t(10) = 3.13$, $p < .011$, $d = .94$. It is likely that this result was not statistically significant as a result of the low number of observations and hence low statistical power.

Contrary to *a priori* hypotheses, the accuracy of estimates did not appear to be a function of the frequency or duration of the behavior relative to other symptoms. Nevertheless, post hoc exploratory analyses were conducted to further test this hypothesis. Specifically, a series of one-way ANOVAs were run to test whether the difference between participants' correlation estimations and EMA correlations varied as a function of the absolute value of participants' behavioral frequency and duration. To accomplish this, difference scores were calculated by subtracting the EMA correlations from the correlation estimates. These differences were then ranked according to the absolute value of participants' Time 3 recalled frequency and duration. Participants were then grouped into tertiles representing low, medium, and high frequency and duration. Mean absolute total weekly frequencies for each tertile were as follows: Low $M = 1.83$, Range 0 to 7; Medium $M = 18.02$, Range 7.5 to 30; and High $M = 601.21$, Range 35 to 15000. Mean absolute daily duration in minutes for each tertile were as follows: Low $M = 0.72$, Range 0 to 2; Medium $M = 13.99$, Range 2.5 to 30; and High $M = 159.80$, Range 45 to 720.

Given the exploratory nature of these analyses, the significance level was set at .05 despite multiple comparisons. For the absolute value of behavioral frequency, differences between estimated and EMA correlations were statistically significant only for the supplemental item sleep, $F(2, 101) = 3.39$, $p = .04$. Differences for all other supplemental items tested (i.e., mood, stress, anxiety, and loneliness) yielded statistically non-significant $F$ values. Thus, a statistically significant difference existed between the Low, Medium, and

High tertiles for the supplemental item sleep, but not for the other supplemental variables. For the absolute value of daily behavioral duration, differences between estimated and EMA correlations were statistically significant for sleep, $F(2, 101) = 5.03, p = .008$; mood $F(2, 115) = 3.29, p = .04$; and loneliness $F(2, 99) = 4.09, p = .02$. Differences for the supplemental items stress and anxiety did not yield significant $F$ values. Post hoc inspection of mean differences revealed that all significant $F$ values resulted because of differences between the Low and High tertiles. Of these, three resulted because the magnitude of overestimation was greater for the High tertile than the Low tertile. Only the ANOVA conducted on behavioral duration of the supplemental item mood revealed the opposite pattern, in which overestimates occurred more in the Low tertile than the High tertile. Equally important, inspection of group means suggests that non-significant results occurred although there were relatively high differences between the estimated and EMA correlations across Low, Medium, and High groups. Thus, these results suggested that differences between estimated and EMA correlations vary as a function of the absolute value of behavioral frequencies and durations. Specifically, these differences were driven by differences between the Low and High tertiles. Some differences were not statistically significant, however, because participants were uniformly inaccurate in estimating correlations regardless of the frequency or duration of the behaviors. Although interesting, these results must be interpreted with caution because a single participant's estimated and EMA correlations contributed to multiple groups, thus violating the assumption of independence of observations. That said, violation of this assumption normally leads to the increase of Type I error. Given the statistically non-significant $F$ values observed, however, this may not have occurred.

Consistent with results from the overall comparisons, these results highlight the exaggerated magnitude of participants' estimated symptom covariation when compared to EMA correlations. With one exception for both behavioral frequencies and durations, all estimated correlations were greater in magnitude than EMA correlations. This suggests that both overall and across the rank order of behavioral frequencies and durations, participants estimated stronger relationships than EMA correlations would support. A paired-samples $t$-test was conducted to test this conclusion. Results indicated that for behavioral frequency, the mean estimated correlation ($M = .17$, $SD = .45$) was statistically significantly greater than mean EMA correlation ($M = .10$, $SD = .30$), $t(585) = 3.58$, $p < .001$. Likewise, results indicated that the mean estimated correlation for duration ($M = .17$, $SD = .46$) was significantly greater than the mean EMA correlation ($M = .10$, $SD = .32$), $t(586) = 3.34$, $p < .001$.

*Discrepancies in Estimated and EMA Correlations for OCD Behaviors Grouped by Subscale.* To examine whether statistically significant differences between participants' estimations and EMA derived calculations occurred more prominently with one type of OCD behavior versus another, data were collapsed across individuals and behavioral rank orders. The means, standard deviations, and test statistics for estimated and EMA correlations grouped according to OCI-R subscales are presented in Table 11 for behavioral frequencies and Table 12 for behavioral durations. Subscale behaviors were examined instead of individual items because the algorithm used to select behaviors for estimation at Time 3 favored the first occurrence of each endorsed behavior per subscale. Thus, inspection of individual behaviors would have resulted in statistics based on one or two individuals for some items.

Similar to calculations based on the rank order of frequencies and durations, comparisons between participants' correlation estimates and EMA correlations consistently resulted in statistically significant differences for the variables anxiety ($d = .29 – 1.16$ for symptom frequency; $d = .16 – 1.11$ for symptom duration) and stress ($d = .31 – .86$ for symptom frequency; $d = .21 – .84$ for symptom duration) across OCI-R subscales. In each comparison, participants overestimated the magnitude of the correlation when compared to the EMA values. This suggests that when estimating covariation, participants attribute more explanatory importance to the variables of stress and anxiety than actually exists.

With respect to the variables sleep, mood, and loneliness, consistent statistical differences were not detected for either behavioral frequencies or durations. However, inspection of effect sizes in Tables 11 and 12 found that moderate and large effect sizes regularly emerged for these variables. Given that subjectively defined interpersonal fights were infrequently encountered by participants, statistics for these variables should be interpreted with caution.

Discussion

The results of this study partially support the contention that patients diagnosed with OCD are inaccurate reporters of their symptoms. Although inaccuracies with respect to retrospective recall of EMA assessed criterion values were largely unsubstantiated, evidence of recall bias was observed in participants' ability to recall functional relations of their symptoms. Results of this study also support the contention that EMA does not result in significant reactivity.

*Hypothesis 1: OCD patients' retrospective recall of OCD symptoms will differ significantly from EMA criterion data*

Results from both paired-samples *t*-tests and random-effects regression analyses failed to detect consistent, statistically significant differences between retrospectively recalled OCD symptoms and EMA criterion data. These largely non-significant results were consistent across analyses for behavioral frequency and duration. The consistency of these results is striking given the use of these two different statistical procedures. Although Type II error may have increased in the paired-samples *t*-tests as a function of the relatively small sample size used for each EMA-OCD item and single error term, these arguments do not apply to the random effects regression.

The practical significance of the observed differences between retrospectively recalled OCD symptoms and EMA criterion data is less consistent across the two analytic approaches. Results from effect size analyses based on the paired samples *t*-tests suggest that when participants were inaccurate, they tended to underestimate the frequency of their behaviors relative to the criterion value, but overestimate the duration. However, the effect sizes are mathematically related to the paired samples *t*-test, and therefore susceptible to

similar concerns regarding statistical conclusion validity. In contrast, the coefficient of recall

in the random-effects analyses provides information about the average difference between

retrospective recall and EMA criterion values after accounting for predictor variables

included in the model and partialing of unexplained variance. As a result, the coefficient of

recall represents a statistically unbiased estimator of population parameters. Results from

these analyses suggest that across the 18 EMA-OCD behaviors, participants' inaccuracies

were of little practical importance. The regression models of average daily frequency suggest

that across the 18 OCD behaviors, differences between retrospectively recalled daily

frequencies and EMA calculated daily frequencies exceeded 15 occurrences per day only

once (i.e., "I repeated certain numbers"). Similarly, average daily inaccuracy in estimated

duration exceeded 15 minutes per day on only two variables (i.e., "I checked something I

didn't need to" and "I repeated certain numbers"). Thus, on the whole, the magnitude of

differences between participants' retrospective recall and EMA criterion data is unlikely to

lead to drastic inaccuracies in clinical case conceptualization.

These results stand in partial contrast to inaccuracies reported in previous studies of

retrospective recall accuracy using EMA designs (Stein & Corte, 2003; Stone et al., 2004).

Several factors may account for these differences. First, participants in this study recorded

and recalled more behaviors. Second, target behaviors in this study were recorded using a 0

to infinity scale for behavioral frequency and a 0 to 240-minute scale for behavioral duration.

In contrast, Stone et al. utilized a 100-point visual analogue scale to record participants' pain

intensity. Third, this study examined the magnitude of differences using *t*-tests and random

effects regression, whereas Stein and Corte examined differences using the nonparametric

Wilcoxen Test. Finally, OCD patients may differ from eating disordered patients and pain

patients in a way that makes them more accurate for this task. Future research is needed to clarify which of these hypotheses is more plausible.

The results of this study also failed to support contentions in the OCD literature of memory inaccuracies, at least as measured in this study. Laboratory research has concentrated on episodic memory of experimental stimuli and found the most consistent evidence of memory inaccuracies for non-verbal stimuli (Muller & Roberts, 2005). Given that this was the first study to examine memory difficulties of OCD patients in an ecologically valid manner, it is unclear whether verbal memory processes, non-verbal memory processes, or both are involved in participants' recalled symptoms. Whereas participants provided answers to the questions in this study in a verbal format, it is nevertheless possible that the actual memories are at least partially non-verbal. Regardless, the present study calls into question the ecological validity of laboratory findings. Future studies utilizing varied stimuli both in the laboratory and via EMA are needed to tease apart these issues.

*Hypothesis 2: Ecological Momentary Assessment of OCD symptoms will not result in statistically significant reactivity*

Results from the current study suggest that reactivity did not occur in response to the EMA recording method. Instead, reactivity in this study was observed during the first week between Time 1 and Time 2. These results add to the growing body of literature that has failed to detect reactivity as a result of EMA (e.g., Cruise et al., 1996; Hufford et al., 2002; Stein & Corte, 2003). The failure to detect reactivity during EMA is important for this study because it eliminates one threat to internal validity, namely that differences noted during recall tasks are confounded by reactivity during EMA.

The current study also demonstrated the importance of including a baseline period prior to the collection of EMA criterion values. Multiple factors likely caused the decline in symptom scores observed from Time 1 to Time 2. Hypothesized factors include increased awareness of actual symptoms, increased familiarity with questions asked during the assessment, instructions to "pay attention to your symptoms," and improvement in symptoms secondary to enrolling in the study and speaking to an assessor. In this study, the sum of these effects resulted in statistically observable reactivity from Time 1 to Time 2, whereas the addition of EMA from Time 2 to Time 3 did not. Failure to include the baseline period would have relegated the decline in symptoms to the focal week for collecting criterion values, thus obfuscating inferences regarding retrospective recall and symptom change.

The statistically significant decreases in symptom severity observed between Time 1 and Time 2 occurred despite patients' overall longstanding length of symptoms and treatment. As indicated above, this suggests that some aspect of the current study's design triggered the reactivity during the first week. It is possible that the subtle instruction to "Pay attention to your symptoms over the next week" triggered this response. Alternately, assessments administered at Time 1 may have cued participants to think more closely about these behaviors than they had prior to the Time 2 assessment. To the degree that this is true, Time 2 assessments may represent more accurate assessments than when participants were completely "naive" at Time 1. Regardless of the cause, the observed reactivity from Time 1 to Time 2 in the absence of programmed treatment changes suggests that traditional assessment in outcome studies may be biased towards overinflating symptom change. Future studies with a control group who do not receive informal instructions at Time 1 may help clarify the cause of the drop in symptoms.

*Hypothesis 3: Participants' estimations of symptom covariation magnitude will be statistically different from covariation statistics derived from EMA data*

This study was the first to examine patients' ability to estimate the magnitude of functional relations between their focal symptoms and supplemental variables. Results suggest that participants consistently overestimated the magnitude of correlations between OCD behaviors and supplemental items. These effects were most robust for the relationship between OCD behaviors and the supplemental items stress and anxiety, though nearly all participants' estimated correlations were overestimates across behavioral frequencies, durations, and OCD symptom subtype.

These results are consistent with O'Brien (1995), who found that advanced graduate students' estimates of covariation based upon hypothetical self-monitoring data were inaccurate. Graduate students overestimated weak correlations and underestimated strong correlations. This study extended O'Brien's findings of the overestimation of weak correlations. Whether OCD patients also underestimate strong correlations, like the graduate students in O'Brien's study, could not be tested because strong correlations calculated between OCD symptoms and supplemental items were not observed.

The impact of overestimated correlations on clinical decision-making is currently unknown. Overestimations of symptom covariation could have a negligible impact on treatment conceptualization, or it could lead clinicians to concentrate on attempting to modify relationships that are largely spurious. Future studies are needed to understand the pervasiveness of the overestimation effect, its consequences, and the degree to which it can be observed in other patient and non-patient populations.

Although ascertaining magnitudes of EMA correlations was not a main aim of this study, these data represent one of the only real-time data sets collected on OCD symptoms and therefore provide important insights into the dynamics of OCD symptoms. Inspection of these data suggest relatively low correlations between OCD symptoms and the supplemental variables. The low magnitude of EMA correlations is interesting in its own right as it seems to contradict theories that posit an increase in the duration or frequency of symptoms in the presence of stress and anxiety. By way of extension, these findings suggest that variation in OCD symptomatology is dependent either on variables not assessed in this study or that multifactorial models are needed to predict symptom covariation.

*General implications of findings*

Overall, this study suggests that OCD patients accurately recall the frequency and duration of their OCD symptoms following one week of EMA. Despite accuracy in retrospective recall tasks, OCD patients overestimated their symptom covariation. OCD patients displayed greater accuracy than other patient populations tested using a similar EMA design (Stein & Corte, 2003; Stone et al., 2004), although estimates of symptom covariation were largely inaccurate. The importance of this finding lies in the following question: If patients who are queried four times per day for a week demonstrate the ability to accurately recall symptoms and yet overestimate symptom covariation, how inaccurate are patients' reports of symptom covariation when given to treatment providers?

If the overestimation effect is replicated in other patient samples, researchers will be called upon to explain why it occurs. It is possible that the cognitive load required to consciously attend to factors that covary with symptoms exceeds the capacity of humans. Proponents of statistical decision-making would likely endorse such a position (e.g., Garb,

1998). It is also possible that if patients were taught the concept of correlations prior to self-monitoring and told they will be asked to report on their symptom covariation, they may be able to produce more accurate covariation estimates. The fact remains, however, that this is not and never will be part of standard clinical practice. Therefore, in therapeutic settings, clinicians may base case conceptualizations on inaccurate data.

*General limitations of study*

The study was limited in several ways. First, although the sample size in this study was larger than a similarly designed study that found statistically significant retrospective recall differences ($n = 35$ vs. $n = 16$), it was still relatively small. A greater sample size may have increased the chances of detecting statistically significant differences in the retrospective recall task. However, if the magnitude of differences observed in this sample replicate, differences between retrospective recall and EMA criterion data lack practical significance regardless of the results of statistical tests. That said, because a number of OCD symptoms were not endorsed by all participants, the actual sample sizes for $t$-tests differed across symptoms and were lower than the total N of 35 participants. Only after data collection commenced was it clear that the idiosyncratic nature of OCD would lead most participants to endorse only a portion of the EMA-OCD questions and that non-endorsed questions would nearly always lead to perfect, though theoretically less interesting, retrospective recall.

In this study, each participant functioned as a control for him/herself. Nevertheless, a second limitation of this study is the lack of a control group. Although we fully acknowledge that a control group could have added to the strength of the study, a separate control group would have been needed for each hypothesis. Furthermore, a control group would have been

very difficult to generate for this study. For instance, conclusions regarding hypotheses of retrospective recall and correlation estimations would have been strengthened by inclusion of non-OCD patients. However, such a control group would have necessitated alternate items relevant to those participants or led to complete zero-order responses because the participants do not suffer from OCD. Either situation would have compromised the validity of the control condition. In contrast, a control group would have been possible with respect to the examination of reactivity. However, given the base rate of OCD and resultant length of participant recruitment, the control group did not seem justified.

*Future directions*

Results of this study suggest several directions for continued inquiry. Future studies are needed to clarify why data from this study stands in contrast to results from previous investigations of retrospective recall accuracy using similar designs (Stein & Corte, 2003; Stone et al., 2004). Of primary concern is whether this study achieved divergent results as a function of the questions asked, or if OCD patients retrospectively recall their symptoms better than other patients. If results from the current study are robust, then future research is also needed to further test the external validity of laboratory studies that have detected inaccuracies in OCD patients on various measures of memory.

Given that this is the first study to examine patients' ability to estimate symptom covariation, little is understood about this phenomenon. Further studies are needed to understand the implications of symptom overestimation and the factors that mediate covariance estimation errors. Furthermore, the supplemental items utilized in the covariation estimation tasks represent only a fraction of the multiple theoretical variables currently hypothesized to maintain OCD symptomatology. Inclusion of such variables in future studies

would test the generalizability of both the present findings and hypothesized theories generated using cross-sectional data.

Relatively little is known about processes patients used to generate retrospective recall of behavioral frequencies, durations, and symptom covariations. Future studies might address this issue by identifying low and high frequency behaviors, then querying patients as to whether they utilized estimation or enumeration strategies as a function of behavioral frequency or duration. Informal observation during data collection for this study suggests the use of estimation procedures for high frequency behaviors. For instance, thinking out loud during the recall task, one participant stated, "I answered x times every time the PDA beeped. It beeped four times per day. That makes x occurrences every day for seven days. That means I must have done it y times."

Analyses in this study also suggest that further research is needed to understand the daily fluctuations of OCD symptomatology. As suggested previously, either different variables or multifactoral models will likely be needed to better understand variations in OCD symptomatology. The high response rate of participants in this study over a relatively time-consuming protocol suggests OCD patients can adhere to an EMA protocol. In fact, some patients said they did not want to give back the PDA's because they felt it helped them understand themselves better.

In addition to future studies that necessitate data collection, additional analyses on these data may provide further insights into the dynamics of OCD patients' retrospective recall and symptom covariation estimates. For instance, analyses in this study concentrated on attributes of central tendency. Given that some participants were more accurate than others, it is possible that symptom variability predicts accuracy. Likewise, symptom

variability may also predict those participants who can more accurately estimate symptom covaration.

Data collected during this study may also facilitate the psychometric investigations of how EMA data correspond to other assessment modalities such as clinician ratings (i.e., ADIS-IV and Y-BOCS) and self-report questionnaires (i.e., OCI-R and BDI-II). For instance, EMA frequency and duration data can be compared against the Likert response formats of the OCI-R to determine what constitutes "often," "seldom," and so on for participants. Studies of human judgment suggest that the subjective context in which a participant answers questions can have profound effects on their ratings (Birnbaum, 1999). Thus, the three assessment time points plus the week of EMA observation may provide clues as to the subjective contexts participants experienced during the study.

In conclusion, it is presently unclear what consequences result from the effects observed in this study. Although it remains an empirical question, it is possible that accuracy is unimportant in therapy, where a therapist and patient construct "truth" together over time (Back, 1994). However, ascertainment of accurate data is a prerequisite to make valid inferences and scientific progress.

Table 1.

*Characteristics of studies designed to compare retrospective and in vivo assessments*

| Study | Topic | Sample | Length of recording | Format | Type of Assessment/ Frequency of Signals | N | Results | Cohen's d |
|---|---|---|---|---|---|---|---|---|
| de Beurs et al., 1992 | Frequency of panic attacks | Agoraphobics | 12 weeks | Paper-and-pencil | Event contingent | 32 | Twice as many retrospectively reported panic attacks at pretest compared to self-monitoring reports from *following* week. Improvement of correspondence between data collection procedures over the course of the study. Only comparison of same time frame (week 6) yielded less discrepancy. | -0.16 (at week 6 of the study) |
| Frederiks en et al., 1975 | Reliability of different self-monitoring schedules (smoking frequency) | Student smokers | 1 week | Paper-and pencil | Multiple schedules | 15 | Continuous recording more accurate relative to daily or weekly recording. However, all procedures resulted in at least 85% agreement with objective criteria | Relevant data not reported |
| Marks et al., 1999 | Frequency of cognitions and physiologica l symptoms during | Agoraphobics | 24 hours | Paper-and-pencil | Event contingent – during an exposure exercise | 20 | Retrospective recall of cognitions grater than in vivo data especially for catastrophic thoughts about physical symptoms. However, no discrepancy | Relevant data not reported |

| Study | Topic | Sample | Length of recording | Format | Type of Assessment/ Frequency of Signals | N | Results | Cohen's d |
|---|---|---|---|---|---|---|---|---|
| | exposure | | | | | | between retrospective recall and in vivo data of physiological symptoms. | |
| Parkinson et al., 1995 | Positive & Negative Mood assessed on a 20- point visual analog scale | Non-clinical adults | 2 weeks | Handheld Computer | Signal-Contingent/ Every 2 hours | 30 | Daily retrospective reports significantly higher than momentary reports for positive, but not negative mood. Similarly, weekly retrospective reports significantly higher than daily retrospective reports for positive, but not negative mood. | *SD* not reported for group means |
| Shiffman et al., 1997 | Episodes of smoking lapses and temptations | Smokers | 2 week baseline plus time to lapse (up to 25 days)/ retrospecti ve recall occurred 12 weeks later | Handheld computer | Event and Signal Contingent/ Events + 4-5 random assessments | 127 | Recall of EMA-recorded smoking lapses poor across items measuring mood, activity, episode triggers, and abstinence violation effects. Average kappas of these domains ranged from 0.18 to 0.27. | Relevant data not provided |

| Study | Topic | Sample | Length of recording | Format | Type of Assessment/ Frequency of Signals | N | Results | Cohen's d |
|-------|-------|--------|---------------------|--------|------------------------------------------|---|---------|-----------|
| Stein et. al, 2003 | Eating disorder behaviors | Women with threshold or subthreshold anorexia and bulimia nervosa | 4 weeks | Handheld Computer | Event Contingent Sampling | 16 | EMA-recorded binging, excessive exercise, and sum of all eating disorder behaviors significantly lower than retrospective recall. Not all estimates of behaviors compared between retrospective assessment and EMA were significantly different. | +0.44 (binging), +0.88 (excessive exercise), & +0.51 (across all eating disorder behaviors) |
| Stone et al., 1998 | Engagement in coping behaviors, assessed dichotomously | Community sample with job or marital stress | 2 days | Handheld Computer | Signal Contingent/ on average every 40 minutes | 100 | Compared to EMA data, on average 29% underreported and 21% overreported coping on the retrospective recall. Overall, 23% discrepancy between two methods. Underreporting more likely at retrospective recall with cognitive coping, overreporting more likely with behavioral coping. | Data not reported |
| Stone et al., 2004 | Pain intensity, assessed on a 100 pt visual analogue scale | Chronic pain patients | 2 one-week reference periods | Handheld Computer | Signal contingent | 68 | Correspondence between retrospective recall and in vivo data moderate to high for between-person analysis, but low for with person analysis. | + .93 (1st week); +.97 (2nd week). |

*Note.* Relative to the criterion data, negative effect sizes represent underestimates and positive effect sizes represent overestimates of retrospective assessment.

Table 2.

*EMA and Retrospective Recall of Total Behavioral Frequencies Across One Week*

| Item | n | EMA | Recall | t | p | d |
|---|---|---|---|---|---|---|
| 1. I saved or collected something I don't need | 21 | 9.52 (17.08) | 5.86 (7.53) | -1.15 | .263 | -.25 |
| 2. I checked something I didn't need to | 23 | 57.43 (102.75) | 17.11 (15.19) | -1.87 | .075 | -.39 |
| 3. I got upset if things were not arranged properly | 21 | 32.67 (52.83) | 13.07 (20.85) | -2.36 | .027 | -.52 |
| 4. I found myself repeatedly counting objects | 16 | 84.69 (163.15) | 85.38 (247.76) | 0.02 | .988 | .00 |
| 5. I round myself not wanting to touch an object touched by someone else | 18 | 20.22 (32.24) | 15.47 (28.18) | -1.67 | .114 | -.39 |
| 6. I found it difficult to control my own thoughts | 30 | 611.40 (2751.34) | 592.23 (2725.80) | -0.67 | .512 | -.12 |
| 7. I collected things I don't need | 17 | 6.59 (7.44) | 10.26 (16.13) | 1.19 | .253 | .29 |
| 8. I repeatedly checked doors, windows, drawers, etc. | 17 | 26.59 (45.51) | 28.32 (71.46) | 0.25 | .809 | .06 |
| 9. I got upset when someone changed the way I arranged things | 19 | 9.16 (12.86) | 6.18 (8.56) | -1.83 | .083 | -.42 |
| 10. I repeated certain numbers | 13 | 136.69 (259.93) | 1267.92 (4135.08) | 1.05 | .315 | .29 |
| 11. I washed or cleaned myself because I felt contaminated | 21 | 26.43 (40.21) | 27.69 (63.75) | 0.20 | .843 | .04 |
| 12. Unpleasant thoughts came into my mind against my will | 27 | 620.85 (2591.86) | 628.19 (2874.62) | 0.12 | .905 | .02 |
| 13. I avoided throwing something away because I was afraid I might need it later | 19 | 6.84 (10.03) | 8.66 (11.48) | 0.66 | .517 | .15 |
| 14. I checked stove burners, water taps, and/or light switches after turning them off | 15 | 13.13 (17.63) | 20.50 (45.90) | 0.71 | .488 | .18 |
| 15. I arranged things in a particular order | 19 | 45.32 (56.46) | 15.61 (18.00) | -2.56 | .020 | -.59 |
| 16. I thought about good and bad numbers. | 11 | 97.27 | 58.00 | -1.14 | .281 | -.34 |

| | | (139.42) | (80.72) | | | |
|---|---|---|---|---|---|---|
| 17. I washed my hands longer than necessary | 13 | 11.00 | 6.58 | -1.05 | .315 | -.29 |
| | | (15.29) | (5.94) | | | |
| 18. I had troubling thoughts that were difficult to get rid of | 30 | 598.17 | 606.75 | 0.20 | .844 | .04 |
| | | (2724.65) | (2729.01) | | | |

*Note.* Negative numbers represent underestimates of the retrospective recall relative to the EMA criterion value and positive numbers overestimates.

Table 3.

*EMA and Retrospective Recall of Average Daily Duration in Minutes per Behavior Across One Week*

| Item | n | EMA | Recall | t | p | d |
|---|---|---|---|---|---|---|
| 1. I saved or collected something I don't need | 21 | 5.57 | 3.83 | -0.71 | .489 | -.15 |
| | | (10.17) | (6.96) | | | |
| 2. I checked something I didn't need to | 23 | 24.94 | 24.11 | -0.14 | .890 | -.03 |
| | | (40.34) | (35.47) | | | |
| 3. I got upset if things were not arranged properly | 21 | 12.95 | 12.17 | -0.26 | .795 | -.06 |
| | | (16.02) | (14.89) | | | |
| 4. I found myself repeatedly counting objects | 16 | 14.46 | 54.25 | 1.54 | .145 | .39 |
| | | (23.95) | (124.88) | | | |
| 5. I round myself not wanting to touch an object touched by someone else | 18 | 12.98 | 8.24 | -1.18 | .254 | -.28 |
| | | (26.12) | (14.84) | | | |
| 6. I found it difficult to control my own thoughts | 30 | 92.56 | 91.46 | -0.06 | .950 | -.01 |
| | | (148.55) | (151.97) | | | |
| 7. I collected things I don't need | 17 | 4.03 | 8.28 | 1.11 | .283 | .27 |
| | | (5.97) | (14.74) | | | |
| 8. I repeatedly checked doors, windows, drawers, etc. | 17 | 7.40 | 13.05 | 1.86 | .081 | .45 |
| | | (10.61) | (15.90) | | | |
| 9. I got upset when someone changed the way I arranged things | 19 | 9.18 | 7.58 | -0.91 | .375 | -.21 |
| | | (16.93) | (13.82) | | | |
| 10. I repeated certain numbers | 13 | 24.21 | 55.02 | 1.80 | .096 | .50 |
| | | (30.35) | (79.30) | | | |
| 11. I washed or cleaned myself because I felt contaminated | 21 | 6.69 | 7.24 | 0.36 | .721 | .08 |
| | | (6.95) | (9.89) | | | |
| 12. Unpleasant thoughts came into my mind against my will | 27 | 103.26 | 108.94 | 0.28 | .781 | .05 |
| | | (163.70) | (163.74) | | | |
| 13. I avoided throwing something away because I was afraid I might need it later | 19 | 6.88 | 8.26 | 0.31 | .759 | .07 |
| | | (15.57) | (15.18) | | | |
| 14. I checked stove burners, water taps, and/or light switches after turning them off | 15 | 3.44 | 10.03 | 2.26 | .041 | .58 |
| | | (4.54) | (11.90) | | | |
| 15. I arranged things in a particular order | 19 | 23.80 | 19.58 | -0.87 | .398 | -.20 |
| | | (28.48) | (30.13) | | | |
| 16. I thought about good and bad numbers. | 11 | 25.95 | 40.39 | 1.07 | .310 | .32 |

|  |  | (29.22) | (53.83) |  |  |  |
|---|---|---|---|---|---|---|
| 17. I washed my hands longer than necessary | 13 | 4.80 | 4.50 | -0.10 | .926 | -.03 |
|  |  | (9.43) | (5.49) |  |  |  |
| 18. I had troubling thoughts that were difficult to get rid of | 30 | 99.00 | 95.11 | -0.16 | .875 | -.03 |
|  |  | (175.99) | (153.67) |  |  |  |

*Note.* To equate metrics with the recall task ("On average, how long did you spend doing … each day?"), EMA duration = weekly duration total/ 7; Negative numbers represent underestimates of the retrospective recall relative to the EMA criterion value and positive numbers overestimates.

Table 4.

*Summary of Random Effects Regression Analysis for Average Daily Behavioral Frequency*

| Item | df | Recall Coefficient | z | p | 95% CI | Overall $R^2$ | Rho |
|---|---|---|---|---|---|---|---|
| 1. I saved or collected something I don't need | 168 | -.60 | -0.70 | 0.484 | -2.28 - 1.08 | .11 | .16 |
| 2. I checked something I didn't need to | 186 | -6.35 | -2.65 | 0.008** | -11.06 - -1.65 | .08 | .66 |
| 3. I got upset if things were not arranged properly | 169 | -3.07 | -1.33 | 0.184 | -7.60 - 1.46 | .10 | .30 |
| 4. I found myself repeatedly counting objects | 125 | -1.96 | -0.26 | 0.791 | -16.50 - 12.58 | .07 | .46 |
| 5. I round myself not wanting to touch an object touched by someone else | 143 | -.81 | -0.81 | 0.420 | -2.77 - 1.15 | .38 | .39 |
| 6. I found it difficult to control my own thoughts | 246 | -12.14 | -0.32 | 0.751 | -87.04 - 62.75 | .22 | .78 |
| 7. I collected things I don't need | 134 | .43 | 0.82 | 0.411 | -.60 - 1.46 | .12 | .25 |
| 8. I repeatedly checked doors, windows, drawers, etc. | 134 | -.12 | -0.12 | 0.908 | -2.22 - 1.97 | .34 | .77 |
| 9. I got upset when someone changed the way I arranged things | 151 | -.54 | -0.78 | 0.437 | -1.91 - .82 | .16 | .11 |
| 10. I repeated certain numbers | 100 | 139.30 | 2.89 | 0.004** | 44.97 - 233.63 | .24 | .09 |
| 11. I washed or cleaned myself because I felt contaminated | 169 | -.41 | -0.32 | 0.747 | -2.94 - 2.11 | .37 | .47 |
| 12. Unpleasant thoughts came into my mind against my will | 222 | -3.67 | -0.08 | 0.936 | -93.42 - 86.09 | .25 | .66 |
| 13. I avoided throwing something away because I was afraid I might need it later | 151 | .06 | 0.12 | 0.908 | -.89 - 1.00 | .22 | .29 |
| 14. I checked stove burners, water taps, and/or light switches after turning them off | 117 | .78 | 0.69 | 0.492 | -1.44 - 3.00 | .21 | .21 |
| 15. I arranged things in a particular order | 151 | -4.74 | -1.78 | 0.075 | -9.95 - .47 | .23 | .09 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 16. I thought about good and bad numbers. | 84 | -7.09 | -1.39 | 0.165 | -17.09 - 2.91 | .51 | .55 |
| 17. I washed my hands longer than necessary | 101 | -.85 | -0.54 | 0.591 | -3.95 - 2.25 | .09 | .01 |
| 18. I had troubling thoughts that were difficult to get rid of | 246 | -8.32 | -0.23 | 0.815 | -77.93 - 61.29 | .22 | .80 |

*Note.* $* = p < .05$; $** = p < .008$

Table 5

*Summary of Random Effects Regression Analysis for Average Daily Behavioral Duration*

| Item | df | Recall Coefficient | z | p | 95% CI | Overall $R^2$ | Rho |
|---|---|---|---|---|---|---|---|
| 1. I saved or collected something I don't need | 168 | -1.65 | -0.29 | 0.772 | -12.87 - 9.56 | .06 | .00 |
| 2. I checked something I didn't need to | 186 | -2.29 | -0.35 | 0.724 | -14.97 - 10.40 | .19 | .66 |
| 3. I got upset if things were not arranged properly | 169 | -2.12 | -0.34 | 0.735 | -14.41 - 10.17 | .21 | .12 |
| 4. I found myself repeatedly counting objects | 125 | 36.08 | 2.88 | 0.004** | 11.51 - 60.65 | .17 | .46 |
| 5. I round myself not wanting to touch an object touched by someone else | 143 | -5.94 | -0.56 | 0.575 | -26.70 - 14.82 | .19 | .19 |
| 6. I found it difficult to control my own thoughts | 246 | -9.95 | -0.53 | 0.599 | -47.00 - 27.10 | .20 | .69 |
| 7. I collected things I don't need | 134 | 4.42 | 1.22 | 0.223 | -2.69 - 11.54 | .09 | .00 |
| 8. I repeatedly checked doors, windows, drawers, etc. | 134 | 4.01 | 0.98 | 0.329 | -4.04 - 12.07 | .11 | .49 |
| 9. I got upset when someone changed the way I arranged things | 151 | -1.80 | -0.43 | 0.669 | -10.07 - 6.46 | .21 | .43 |
| 10. I repeated certain numbers | 100 | 28.596 | 2.44 | 0.015* | 5.66 - 51.53 | .21 | .61 |
| 11. I washed or cleaned myself because I felt contaminated | 169 | -.50 | -0.18 | 0.858 | -6.02 - 5.02 | .16 | .23 |
| 12. Unpleasant thoughts came into my mind against my will | 222 | -2.30 | -0.09 | 0.930 | -53.51 - 48.91 | .21 | .59 |
| 13. I avoided throwing something away because I was afraid I might need it later | 151 | -4.77 | -0.19 | 0.849 | -53.72 - 44.18 | .08 | .00 |
| 14. I checked stove burners, water taps, and/or light switches after turning them off | 117 | 5.52 | 1.61 | 0.107 | -1.19 - 12.24 | .12 | .11 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 15. I arranged things in a particular order | 151 | -4.91 | -0.62 | 0.535 | -20.45 - 10.62 | .38 | .13 |
| 16. I thought about good and bad numbers. | 84 | 12.66 | 1.08 | 0.280 | -10.33 - 35.64 | .37 | .49 |
| 17. I washed my hands longer than necessary | 101 | -.80 | -0.12 | 0.908 | -14.38 - 12.77 | .10 | .00 |
| 18. I had troubling thoughts that were difficult to get rid of | 246 | -11.82 | -0.34 | 0.731 | -79.29 - 55.64 | .11 | .45 |

*Note.* $* = p < .05$; $** = p < .008$

Table 6.

*Sample Mean and Standard Deviation of Measures at Times 1, 2, and 3*

| Measure | Time 1 | Time 2 | Time 3 |
|---|---|---|---|
| OCI-R Total | 25.80 | 21.04 | 19.97 |
| | (9.38) | (9.81) | (11.10) |
| Washing | 3.37 | 2.69 | 2.29 |
| | (3.57) | (2.89) | (2.59) |
| Obsessing | 6.71 | 5.57 | 5.86 |
| | (3.56) | (3.63) | (3.85) |
| Hoarding | 4.26 | 2.83 | 2.83 |
| | (3.94) | (3.62) | (3.16) |
| Ordering | 4.46 | 4.27 | 3.74 |
| | (3.39) | (3.54) | (3.78) |
| Checking | 3.89 | 2.91 | 2.83 |
| | (3.18) | (3.32) | (3.11) |
| Neutralizing | 3.11 | 2.77 | 2.43 |
| | (3.45) | (3.82) | (3.93) |
| | | | |
| Y-BOCS Total | 22.10 | 22.20 | 20.91 |
| | (3.93) | (5.01) | (4.57) |
| Obsession | 11.23 | 11.01 | 10.13 |
| | (2.22) | (3.02) | (2.62) |
| Compulsion | 10.87 | 11.01 | 10.79 |
| | (2.38) | (2.47) | (2.60) |
| | | | |
| BDI-II Total | 22.00 | 18.23 | 17.69 |
| | (13.61) | (13.92) | (13.86) |

Table 7.

*Significance Level of Polynomial Contrasts*

| Measure | Time 1 and 3 | Time 1 and 2 | Time 2 and 3 |
|---|---|---|---|
| OCI-R | .001** | .001** | .260 |
| Washing | .004* | .057 | .075 |
| Obsessing | .015* | .001** | .422 |
| Hoarding | .001* | .001** | 1.000 |
| Ordering | .163 | .612 | .125 |
| Checking | .008* | .012* | .791 |
| Neutralizing | .124 | .310 | .110 |
|  |  |  |  |
| Y-BOCS Total | .082 | .871 | .024* |
| Obsession | .006* | .646 | .017* |
| Compulsion | .823 | .640 | .437 |
|  |  |  |  |
| BDI-II Total | .001** | .001** | .550 |

*Note.* * = Statistically significant at the .05 level; ** = Statistically significant at the .005 level

Table 8.

*Overall Estimated and EMA Correlations between Frequency and Duration of OCD Behaviors with Supplemental Variables*

| Item | $n$ | Estimated | EMA | $t$ | $p$ | $d$ |
|---|---|---|---|---|---|---|
| Behavioral Frequency | | | | | | |
| Sleep | 104 | -.14 (.31) | -.08 (.23) | -1.61 | .111 | -.16 |
| Mood | 118 | -.22 (.39) | -.11 (.24) | -2.68 | .009 | -.25 |
| Stress | 118 | .45 (.33) | .20 (.26) | 7.58 | .001* | .70 |
| Anxiety | 115 | .51 (.36) | .19 (.25) | 9.15 | .001* | .85 |
| Loneliness | 102 | .19 (.32) | .06 (.29) | 3.57 | .001* | .35 |
| Distress following an Interpersonal Fight | 29 | .32 (.37) | .07 (.62) | 1.82 | .079 | .34 |
| Behavioral Duration | | | | | | |
| Sleep | 104 | -.14 (.31) | -.06 (.26) | -1.91 | .059 | -.19 |
| Mood | 118 | -.22 (.39) | -.12 (.28) | -2.29 | .024 | -.21 |
| Stress | 118 | .45 (.33) | .21 (.28) | 7.02 | .001* | .64 |
| Anxiety | 114 | .51 (.36) | .21 (.28) | 7.83 | .001* | .73 |
| Loneliness | 102 | .19 (.32) | .05 (.26) | 4.04 | .001* | .40 |
| Distress following an Interpersonal Fight | 31 | .32 (.37) | -.06 (.65) | 2.88 | .007* | .52 |

Note. $n$ = number of valid comparisons in the sample. $N$s differ across covariation estimates as a function of the number of infrequent behaviors (i.e., participants did not frequently encounter interpersonal fights) and variables that were constant in the EMA data. Constant variables, and therefore incalculable correlation coefficients, resulted due to or invariant responding for a variable across the week (e.g., a participant slept the same number of hours each night); * = $p < .008$.

Table 9.

*Estimated and EMA Correlations between Frequency of OCD Behaviors with Supplemental Variables Grouped by Rank Order of Frequency*

| Item | $n$ | Estimated | EMA | $t$ | $p$ | $d$ |
|---|---|---|---|---|---|---|
| **Most Frequent Behaviors** | | | | | | |
| Sleep | 43 | -.16 (.37) | -.05 (.26) | -1.62 | .114 | -.25 |
| Mood | 47 | -.23 (.41) | -.18 (.26) | -.83 | .410 | -.12 |
| Stress | 47 | .51 (.31) | .28 (.23) | 5.24 | .001* | .76 |
| Anxiety | 46 | .62 (.28) | .25 (.22) | 8.47 | .001* | 1.25 |
| Loneliness | 40 | .22 (.38) | .13 (.29) | 1.50 | .141 | .24 |
| Distress following an Interpersonal Fight | 13 | .34 (.35) | .11 (.74) | .96 | .356 | .27 |
| **Second Most Frequent Behaviors** | | | | | | |
| Sleep | 28 | -.17 (.32) | -.14 (.20) | -.53 | .600 | -.10 |
| Mood | 31 | -.25 (.43) | -.15 (.22) | -1.16 | .256 | -.21 |
| Stress | 31 | .50 (.39) | .23 (.28) | 3.33 | .002* | .60 |
| Anxiety | 29 | .49 (.43) | .24 (.25) | 2.99 | .006* | .56 |
| Loneliness | 26 | .26 (.33) | .06 (.35) | 2.74 | .011 | .54 |
| Distress following an Interpersonal Fight | 9 | .24 (.45) | -.03 (.51) | 1.17 | .276 | .39 |
| **Third Most Frequent Behaviors** | | | | | | |
| Sleep | 21 | -.10 (.25) | -.04 (.24) | -.77 | .450 | -.17 |
| Mood | 24 | -.11 (.38) | -.00 (.22) | -1.13 | .271 | -.23 |
| Stress | 24 | .38 (.32) | .08 (.26) | 3.67 | .001* | .74 |
| Anxiety | 24 | .47 (.38) | .06 (.24) | 4.50 | .001* | .91 |
| Loneliness | 20 | .15 (.23) | -.04 (.28) | 2.25 | .037 | .50 |
| Distress following an Interpersonal Fight | 6 | .43 (.35) | .07 (.59) | 1.16 | .300 | .47 |
| **Fourth Most Frequent Behaviors** | | | | | | |
| Sleep | 9 | -.08 (.16) | -.11 (.18) | -1.91 | .706 | .13 |
| Mood | 11 | -.35 (.26) | -.01 (.21) | -3.13 | .011 | .94 |
| Stress | 11 | .34 (.25) | .09 (.21) | 7.02 | .011 | .94 |
| Anxiety | 11 | .30 (.28) | .08 (.26) | 7.83 | .023 | .81 |
| Loneliness | 11 | .04 (.12) | .02 (.20) | 4.04 | .758 | .10 |
| Distress following an Interpersonal Fight | N/A | | | | | |

Note. $n$ = number of valid comparisons in the sample. $N$s differ across covariation estimates as a function of the number of infrequent behaviors (i.e., participants did not frequently encounter interpersonal fights) and variables that were constant in the EMA data. Constant variables, and therefore incalculable correlation coefficients, resulted due to invariant responding for a variable across the week (e.g., a participant slept the same number of hours each night); [*] $= p < .008$.

Table 10.

*Estimated and EMA Correlations between Duration of OCD Behaviors with Supplemental Variables Grouped by Rank Order of Duration*

| Item | *n* | Estimated | EMA | *t* | *P* | *D* |
|---|---|---|---|---|---|---|
| **Behaviors with Longest Duration** | | | | | | |
| Sleep | 44 | -.21 (.38) | -.04 (.31) | -2.20 | .033 | -.33 |
| Mood | 47 | -.22 (.47) | -.22 (.32) | .15 | .884 | .02 |
| Stress | 47 | .55 (.35) | .29 (.30) | 4.48 | .001* | .65 |
| Anxiety | 43 | .67 (.32) | .32 (.29) | 6.15 | .001* | .94 |
| Loneliness | 38 | .33 (.39) | .10 (.27) | 3.55 | .001* | .58 |
| Distress following an Interpersonal Fight | 14 | .49 (.40) | .00 (.71) | 2.23 | .044 | .60 |
| | | | | | | |
| **Behaviors with Second Longest Duration** | | | | | | |
| Sleep | 27 | -.11 (.29) | -.08 (.23) | -.58 | .567 | -.11 |
| Mood | 33 | -.26 (.30) | -.10 (.24) | -3.23 | .003* | -.56 |
| Stress | 33 | .42 (.31) | .18 (.26) | 3.53 | .001* | .61 |
| Anxiety | 33 | .41 (.34) | .19 (.24) | 3.20 | .003* | .56 |
| Loneliness | 29 | .10 (.28) | .04 (.25) | 1.21 | .238 | .22 |
| Distress following an Interpersonal Fight | 10 | .30 (.24) | -.27 (.62) | 2.49 | .034 | .79 |
| | | | | | | |
| **Behaviors with Third Longest Duration** | | | | | | |
| Sleep | 21 | -.07 (.25) | -.05 (.19) | -.36 | .724 | -.08 |
| Mood | 22 | -.14 (.41) | -.05 (.25) | -.75 | .460 | -.16 |
| Stress | 22 | .45 (.29) | .11 (.31) | 4.03 | .001* | .86 |
| Anxiety | 22 | .49 (.35) | .07 (.28) | 4.28 | .001* | .91 |
| Loneliness | 20 | .15 (.21) | .01 (.30) | 1.73 | .101 | .39 |
| Distress following an Interpersonal Fight | 6 | .05 (.43) | .07 (.61) | -.07 | .947 | -.03 |
| | | | | | | |
| **Behaviors with Fourth Longest Duration** | | | | | | |
| Sleep | 10 | -.06 (.16) | -.07 (.24) | -1.91 | .059 | .05 |
| Mood | 11 | -.25 (.31) | -.04 (.20) | -1.90 | .086 | -.57 |
| Stress | 11 | .26 (.35) | .12 (.15) | 7.02 | .001* | .33 |
| Anxiety | 11 | .28 (.35) | .12 (.20) | 7.83 | .001* | .41 |
| Loneliness | 10 | .04 (.13) | -.02 (.25) | 4.04 | .001* | .22 |
| Distress following an Interpersonal Fight | 1 | NA | | | | |

Note. $n$ = number of valid comparisons in the sample. $N$s differ across covaration estimates as a function of the number of infrequent behaviors (i.e., participants did not frequently encounter interpersonal fights) and variables that were constant in the EMA data. Constant variables, and therefore incalculable correlation coefficients, resulted due to or invariant responding for a variable across the week (e.g., a participant slept the same number of hours each night); $^* = p < .008$.

Table 11.

*Estimated and EMA Correlations between Frequency of OCD Behaviors with Supplemental Variables Grouped by OCI-R Subscale*

| Item | $n$ | Estimated | EMA | $t$ | $p$ | $D$ |
|---|---|---|---|---|---|---|
| **Washing** | | | | | | |
| Sleep | 10 | -.06 (.11) | -.01 (.12) | -1.62 | .140 | -.51 |
| Mood | 13 | -.22 (.22) | .03 (.19) | -3.13 | .009 | -.87 |
| Stress | 13 | .30 (.29) | -.00 (.30) | 2.81 | .016 | .78 |
| Anxiety | 13 | .36 (.37) | -.02 (.28) | 3.58 | .004[*] | .99 |
| Loneliness | 11 | -.00 (.25) | -.14 (.23) | 1.79 | .104 | .54 |
| Distress following an Interpersonal Fight | NA | | | | | |
| | | | | | | |
| **Obsessing** | | | | | | |
| Sleep | 36 | -.26 (.40) | -.09 (.22) | -2.16 | .038 | -.36 |
| Mood | 39 | -.29 (.47) | -.24 (.25) | -.69 | .494 | .11 |
| Stress | 39 | .58 (.28) | .33 (.22) | 4.86 | .001[*] | .78 |
| Anxiety | 39 | .63 (.28) | .30 (.20) | 7.26 | .001[*] | 1.16 |
| Loneliness | 33 | .38 (.33) | .20 (.29) | 3.06 | .005[*] | .53 |
| Distress following an Interpersonal Fight | 11 | .35 (.23) | .26 (.64) | .47 | .651 | .14 |
| | | | | | | |
| **Hoarding** | | | | | | |
| Sleep | 13 | -.01 (.03) | -.12 (.19) | 2.21 | .047 | .61 |
| Mood | 14 | -.02 (.32) | -.03 (.20) | .10 | .918 | .03 |
| Stress | 14 | .24 (.33) | .03 (.16) | 1.79 | .097 | .48 |
| Anxiety | 12 | .30 (.34) | .08 (.16) | 1.61 | .137 | .46 |
| Loneliness | 11 | .10 (.20) | .22 (.18) | -1.81 | .100 | -.55 |
| Distress following an Interpersonal Fight | 2 | .20 (.71) | -.39 (.16) | -1.51 | .372 | -1.07 |
| | | | | | | |
| **Ordering** | | | | | | |
| Sleep | 15 | -.16 (.34) | -.12 (.21) | -.27 | .786 | -.07 |
| Mood | 17 | -.15 (.50) | -.12 (.21) | -.23 | .825 | -.05 |
| Stress | 17 | .54 (.38) | .20 (.24) | 3.55 | .003[*] | .86 |
| Anxiety | 17 | .57 (.40) | .20 (.24) | 3.33 | .004[*] | .81 |
| Loneliness | 15 | .15 (.26) | -.02 (.28) | 1.84 | .087 | .48 |

| | n | | | | | |
|---|---|---|---|---|---|---|
| Distress following an Interpersonal Fight | 7 | .45 (.33) | -.10 (.51) | 2.09 | .082 | .79 |
| **Checking** | | | | | | |
| Sleep | 19 | -.05 (.30) | -.09 (.30) | .50 | .625 | .11 |
| Mood | 22 | -.32 (.28) | -.05 (.23) | -3.71 | .001* | -.79 |
| Stress | 22 | .49 (.33) | .19 (.28) | 3.46 | .002* | .74 |
| Anxiety | 22 | .56 (.31) | .17 (.28) | 5.19 | .001* | 1.11 |
| Loneliness | 20 | .13 (.36) | -.08 (.29) | 1.96 | .065 | .44 |
| Distress following an Interpersonal Fight | 5 | .48 (.51) | .11 (.80) | .782 | .478 | .35 |
| **Neutralizing** | | | | | | |
| Sleep | 11 | -.12 (.15) | -.03 (.26) | -1.93 | .082 | -.58 |
| Mood | 13 | -.11 (.28) | -.04 (.24) | -.84 | .417 | -.23 |
| Stress | 13 | .28 (.27) | .21 (.18) | 1.13 | .283 | .31 |
| Anxiety | 12 | .29 (.44) | .16 (.18) | .99 | .343 | .29 |
| Loneliness | 12 | .05 (.23) | .08 (.15) | -.17 | .867 | -.05 |
| Distress following an Interpersonal Fight | 4 | -.35 (.51) | .54 (.27) | 1.87 | .158 | .93 |

Note. $n$ = number of valid comparisons in the sample. $N$s differ across covariation estimates as a function of the number of infrequent behaviors (i.e., participants did not frequently encounter interpersonal fights) and variables that were constant in the EMA data. Constant variables, and therefore incalculable correlation coefficients, resulted due to or invariant responding for a variable across the week (e.g., a participant slept the same number of hours each night); * = $p < .008$.

Table 12.

*Estimated and EMA Correlations between Duration of OCD Behaviors with Supplemental Variables Grouped by OCI-R Subscale*

| Item | $n$ | Estimated | EMA | $t$ | $p$ | $d$ |
|---|---|---|---|---|---|---|
| **Washing** | | | | | | |
| Sleep | 10 | -.06 (.11) | .00 (.13) | -1.31 | .223 | -.41 |
| Mood | 13 | -.22 (.22) | .03 (.18) | -3.53 | .004 | -.98 |
| Stress | 13 | .30 (.29) | -.01 (.27) | 2.78 | .017 | .77 |
| Anxiety | 13 | .36 (.37) | -.01 (.27) | 3.40 | .005[*] | .94 |
| Loneliness | 11 | -.01 (.25) | -.11 (.18) | 1.05 | .320 | .32 |
| Distress following an Interpersonal Fight | 0 | NA | | | | |
| | | | | | | |
| **Obsessing** | | | | | | |
| Sleep | 36 | -.26 (.40) | -.07 (.32) | -2.07 | .046 | -.35 |
| Mood | 39 | -.29 (.47) | -.31 (.32) | .18 | .862 | .03 |
| Stress | 39 | .58 (.28) | .36 (.27) | 3.72 | .001[*] | .60 |
| Anxiety | 38 | .63 (.28) | .37 (.26) | 4.69 | .001[*] | .76 |
| Loneliness | 33 | .38 (.33) | .18 (.28) | 3.20 | .003[*] | .56 |
| Distress following an Interpersonal Fight | 12 | .41 (.29) | .17 (.69) | 1.01 | .334 | .29 |
| | | | | | | |
| **Hoarding** | | | | | | |
| Sleep | 13 | .01 (.03) | -.11 (.23) | 1.81 | .095 | .50 |
| Mood | 14 | -.02 (.32) | .00 (.21) | -.19 | .849 | -.05 |
| Stress | 14 | .24 (.33) | .03 (.16) | 1.79 | .097 | .48 |
| Anxiety | 12 | .30 (.34) | .07 (.14) | 1.77 | .104 | .51 |
| Loneliness | 11 | .10 (.20) | .13 (.17) | -.71 | .493 | -.21 |
| Distress following an Interpersonal Fight | 2 | -.20 (.71) | .38 (.17) | -1.53 | .370 | -1.08 |
| | | | | | | |
| **Ordering** | | | | | | |
| Sleep | 15 | -.16 (.34) | -.09 (.20) | -.56 | .587 | -.14 |
| Mood | 17 | -.15 (.50) | -.08 (.25) | -.44 | .668 | -.11 |
| Stress | 17 | .54 (.38) | .17 (.31) | 3.44 | .003[*] | .83 |
| Anxiety | 17 | .57 (.40) | .20 (.27) | 3.28 | .005[*] | .79 |
| Loneliness | 15 | .15 (.26) | -.06 (.25) | 1.99 | .066 | .51 |
| Distress following an Interpersonal Fight | 8 | .40 (.35) | -.18 (.63) | 2.36 | .051 | .83 |

| | n | | | t | p | |
|---|---|---|---|---|---|---|
| **Checking** | | | | | | |
| Sleep | 19 | -.05 (.30) | -.06 (.27) | .08 | .935 | .02 |
| Mood | 22 | -.32 (.28) | -.07 (.17) | -3.71 | .001* | -.79 |
| Stress | 22 | .49 (.33) | .18 (.24) | 3.95 | .001* | .84 |
| Anxiety | 22 | .56 (.31) | .15 (.26) | 5.21 | .001* | 1.11 |
| Loneliness | 20 | .13 (.36) | -.03 (.23) | 1.94 | .067 | .43 |
| Distress following an Interpersonal Fight | 5 | .48 (.51) | -.32 (.68) | 2.16 | .097 | .97 |
| **Neutralizing** | | | | | | |
| Sleep | 11 | -.12 (.15) | .02 (.22) | -2.40 | .037 | -.72 |
| Mood | 13 | -.12 (.29) | -.02 (.29) | -.97 | .351 | -.27 |
| Stress | 13 | .28 (.27) | .22 (.19) | .76 | .463 | .21 |
| Anxiety | 12 | .29 (.44) | .21 (.16) | .56 | .585 | .16 |
| Loneliness | 12 | .04 (.07) | .02 (.26) | .30 | .769 | .09 |
| Distress following an Interpersonal Fight | 4 | .08 (.15) | -.41 (.50) | 2.36 | .100 | 1.18 |

Note. $n$ = number of valid comparisons in the sample. $N$s differ across covariation estimates as a function of the number of infrequent behaviors (i.e., participants did not frequently encounter interpersonal fights) and variables that were constant in the EMA data. Constant variables, and therefore incalculable correlation coefficients, resulted due to or invariant responding for a variable across the week (e.g., a participant slept the same number of hours each night); $^* = p < .008$.
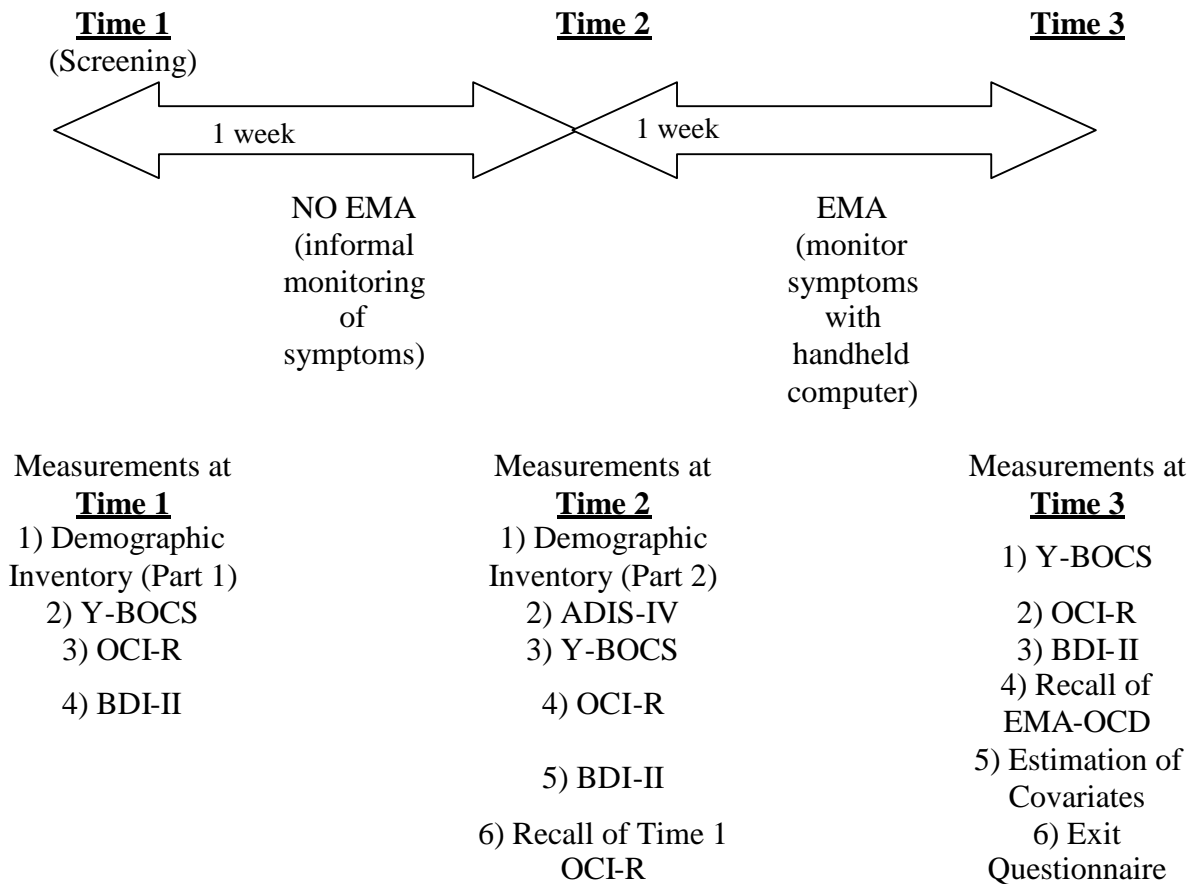
**Time 1**
(Screening)

**Time 2**

**Time 3**

1 week                          1 week

NO EMA
(informal
monitoring
of
symptoms)

EMA
(monitor
symptoms
with
handheld
computer)

Measurements at
**Time 1**
1) Demographic
Inventory (Part 1)
2) Y-BOCS
3) OCI-R

4) BDI-II

Measurements at
**Time 2**
1) Demographic
Inventory (Part 2)
2) ADIS-IV
3) Y-BOCS

4) OCI-R

5) BDI-II

6) Recall of Time 1
OCI-R

Measurements at
**Time 3**

1) Y-BOCS

2) OCI-R
3) BDI-II
4) Recall of
EMA-OCD
5) Estimation of
Covariates
6) Exit
Questionnaire

Figure 1: Study Design

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th, text revision ed.). Washington, DC: Author.

Back, K. W. (1994). Accuracy, truth, and meaning in autobiographical reports. In N. Schwartz & S. Sudman (Eds.), *Autobiographical Memory and the Validity of Retrospective Reports* (pp. 39-53). New York: Springer.

Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic* (2nd ed.). New York: Guilford.

Barton, K. A., Blanchard, E. B., & Veazy, C. (1999). Self-monitoring as an assessment strategy in behavioral medicine. *Psychological Assessment, 11*, 490-497.

Beck, A. T., & Emery, G. (1985). *Anxiety disorders and phobias.* New York: Basic Books.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory Manual* (2nd ed.). San Antonio, TX: Psychological Corporation.

Birnbaum, M. H. (1999). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods, 4*, 243-249.

Borenstein, M., Rothstein, H., & Cohen, J. (1997). Sample Power (Version 1.0): SPSS.

Brewer, W. F. (1994). Autobiographical memory and survey research. In N. Schwartz & S. Sudman (Eds.), *Autobiographical Memory and the Validity of Retrospective Reports* (pp. 11-20). New York: Springer-Verlag.

Brown, T. A., DiNardo, P. A., & Barlow, D. H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV: (Adult Version)*. San Antonio, TX: Psychological Corporation.

Brown, T. A., DiNardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of *DSM-IV* anxiety and mood disorders: Implications for the classification of emotional disorders. *Journal of Abnormal Psychology, 110*(1), 49-58.

Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. *Journal of Personality & Social Psychology, 47*(4), 738-748.

Cruise, C. E., Broderick, J., Porter, L., Kaell, A., & Stone, A. A. (1996). Reactive effects of diary self-assessment in chronic pain patients. *Pain, 67*(2-3), 253-258.

de Beurs, E., Lange, A., & Van Dyck, R. (1992). Self-monitoring of panic attacks and retrospective estimates of panic: Discordant findings. *Behavior Research & Therapy, 30*, 411-413.

Deckersbach, T., Savage, C. R., Curran, T., Bohne, A., Wilhelm, S., Baer, L., Jenike, M. A., & Rauch, S. L. (2002). A study of parallel implicit and explicit information processing in patients with obsessive-compulsive disorder. *American Journal of Psychiatry, 159*(10), 1780-1782.

Dozois, D. J. A., & Dobson, K. S. (2002). Depression. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 259-299). New York: Guilford.

Foa, E. B., Huppert, J. D., Leiberg, S., Langer, R., Kichic, R., Hajcak, G., & Salkovskis, P. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment, 14*, 485-496.

Foa, E. B., Kozak, M. J., Salkovskis, P., Coles, M. E., & Amir, N. (1998). The validation of a new obsessive-compulsive disorder scale: The Obsessive-Compulsive Inventory. *Psychological Assessment, 10*, 206-214.

Frederiksen, L. W., Epstein, L. H., & Kosevsky, B. P. (1975). Reliability and controlling effects of three procedures for self-monitoring smoking. *The Psychological Record, 25*, 255-264.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington DC: American Psychological Association.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Delgato, P., Heninger, G. R., & Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive Scale: II. Validity. *Archives of General Psychiatry, 46*, 1012-1016.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., Heninger, G. R., & Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive Scale: I. Development, use, and reliability. *Archives of General Psychiatry, 46*, 1006-1011.

Haynes, S. N., Leisen, M. B., & Blaine, D. D. (1997). Design of individualized behavioral treatment programs using functional analytic clinical case models. *Psychological Assessment, 9*(4), 334-348.

Hedeker, D. (2004). An introduction to growth modeling. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 215-234). Thousand Oaks, CA: Sage Publications.

Herman, D. J. (1994). The validity of retrospective reports as a function of the directness of retrieval processes. In N. Schwartz & S. Sudman (Eds.), *Autobiographical Memory and the Validity of Retrospective Reports* (pp. 21-37). New York: Springer-Verlag.

Herman, S., & Koran, L. M. (1998). In vivo measurement of obsessive-compulsive disorder symptoms using palmtop computers. *Computers in Human Behavior, 14*(3), 449-462.

Hufford, M. R., Shields, A. L., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. *Psychology of Addictive Behaviors, 16*(3), 205-211.

Karno, M., Golding, J., Sorenson, S., & Burnam, M. A. (1988). The epidemiology of obsessive-compulsive disorder in five U.S. communities. *Archives of General Psychiatry, 45*, 1094-1099.

Kim, S., Dysken, M., & Kuskowski, M. (1990). The Yale-Brown Obsessive Compulsive Scale: A reliability and validity study. *Psychiatry Research, 41*, 37-44.

King, D. W., & King, L. A. (2001). Contemporary approaches to missing data: The glass is really half full. *PTSD Research Quarterly, 12*, 1-5.

Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment, 11*(4), 415-425.

Marks, M., & Hemsley, D. (1999). Retrospective versus prospective self-rating of anxiety symptoms and cognitions. *Journal of Anxiety Disorders, 13*(5), 463-472.

Menon, G. (1994). Judgments of behavioral frequencies: Memory search and retrieval strategies. In N. Schwartz & S. Sudman (Eds.), *Autobiographical Memory and the Validity of Retrospective Reports* (pp. 161-172). New York: Springer.

Menon, G., & Yorkston, E. A. (2000). The use of memory and contextual cues in the formation of behavioral frequency judgments. In A. Stone & J. S. Turkkan & C. A. Bachrach & J. B. Jobe & H. S. Kurtzman & V. S. Cain (Eds.), *The Science of Self-Report: Implications for Research and Practice* (pp. 63-79). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Muller, J., & Roberts, J. E. (2005). Memory and attention in Obsessive-Compulsive

Disorder: A review. *Journal of Anxiety Disorders, 19*, 1-28.

O'Brien, W. H. (1995). Inaccuracies in the estimation of functional relationships using self-

monitoring data. *Journal of Behavior Therapy and Experimental Psychiatry, 26*, 351-

357.

Parkinson, B., Briner, R. B., Reynolds, S., & Totterdell, P. (1995). Time frames for mood:

Relations between momentary and generalized ratings of affect. *Personality and

Social Psychology Bulletin, 21*, 331-339.

Richard, D. C. S., & Haynes, S. N. (2002). Behavioral Assessment, *Encyclopedia of

psychotherapy* (Vol. 1, pp. 165-183): Elsevier.

Rowland, D. L., & Thornton, J. A. (2003). Regression strategies for analyzing the study and

pharmacological treatment of sexual response: ANOVA and beyond. *Annual Review

of Sex Research, 14*, 185-220.

Savage, C. R., Keuthen, N. J., Jenike, M. A., Brown, H. D., Baer, L., Kendrick, A. D.,

Miguel, E. C., Rauch, S. L., & Albert, M. S. (1996). Recall and recognition memory

in obsessive-compulsive disorder. *Journal of Neuropsychiatry, 8*(1), 99-103.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art.

*Psychological Methods, 7*, 147-177.

Shiffman, S. U. (2000). Real-time self-report of momentary states in the natural environment:

Computerized ecological momentary assessment. In A. Stone & J. S. Turkkan & C.

A. Bachrach & J. B. Jobe & H. S. Kurtzman & V. S. Cain (Eds.), *The Science of Self-

Report: Implications for Research and Practice* (pp. 277-296). Mahwah, New Jersey:

Lawrence Erlbaum Associates.

Shiffman, S. U., Hufford, M. R., Hickcox, M., Paty, J., Gnys, M., & Kassel, J. D. (1997). Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of Consulting & Clinical Psychology, 65*(2), 292-300.

Stein, K. F., & Corte, C. M. (2003). Ecologic momentary assessment of eating-disordered behaviors. *International Journal of Eating Disorders, 34*, 349-360.

Stone, A., Broderick, J. E., Shiffman, S. S., & Schwartz, J. E. (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.

Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., Paty, J., Porter, L. S., & Cruise, L. J. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality & Social Psychology, 74*(6), 1670-1680.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine, 16*(3), 199-202.

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Teasdale, J. D., & Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology, 88*, 248-257.

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. Stone & J. S. Turkkan & C. A. Bachrach & J. B. Jobe & H. S. Kurtzman & V. S.

Cain (Eds.), *The Science of Self-Report* (Vol. 29-47). Mahwah, New Jersey:

Lawrence Erlbaum Associates.

Wilhelm, S., McNally, R. J., Baer, L., & Florin, I. (1997). Autobiographical memory in

obsessive-compulsive disorder. *British Journal of Clinical Psychology, 36*(1), 21-31.

Woody, S. R., Steketee, G., & Chambless, D. L. (1995). Reliability and validity of the Yale-

Brown Obsessive-Compulsive Scale. *Behavior Research and Therapy, 33*, 597-605.

Appendix A

Measures

[Demographic Inventory]
[Obsessive-Compulsive Index – Revised]
[Yale-Brown Obsessive-Compulsive Scale]

Demographic Inventory

<u>Administered at Time 1 (Phone Screening)</u>

1. Name
2. Gender
3. Contact Information
    a. Preferred Telephone:
    b. Email:
    c. Address:
4. Year of Birth: _____ (If after 1987, do not continue)
5. Do you have any vision problems? Do you wear glasses? (determine ability to see screen of hand-held)
6. Do you have any difficulty hearing? (determine ability to hear beep of hand-held)
7. Is English your native language?
    a. If no, how well do you understand English?
        i. Very well
        ii. Well
        iii. So-so
        iv. Not very well
        v. Barely understandable

<u>Administered at Time 2 (1<sup>st</sup> Face-to-face interview)</u>

8. Ethnicity
    a. Asian
    b. African-American
    c. White/ Hispanic
    d. White/ Non-Hispanic
    e. Other (specify)
9. What is your highest Degree/ Years of education
10. When were first diagnosed with OCD?
11. When did you first notice symptoms of OCD?
12. Do you take any medications? If yes, please list:
13. How long have you been in therapy? How many sessions? CBT or just meds?
14. What is your current diagnosis?
    a. Are you currently diagnosed with a medical condition?
    b. Are you currently diagnosed with dementia?

Obsessive-Compulsive Inventory – Revised

The following statements refer to experiences that many people have in their everyday lives. Circle the number that best describes **HOW MUCH** that experience has **DISTRESSED or BOTHERED you during the PAST WEEK.** The numbers refer to the following verbal labels:

| 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Not at all | A little | Moderately | A lot | Extremely |

1.  I have saved up so many things that they get in the way.    0 1 2 3 4
2.  I check things more often than necessary.    0 1 2 3 4
3.  I get upset if objects are not arranged properly.    0 1 2 3 4
4.  I feel compelled to count while I am doing things.    0 1 2 3 4
5.  I find it difficult to touch an object when I know it has been touched by strangers or certain people.    0 1 2 3 4
6.  I find it difficult to control my own thoughts.    0 1 2 3 4
7.  I collect things I don't need.    0 1 2 3 4
8.  I repeatedly check doors, windows, drawers, etc.    0 1 2 3 4
9.  I get upset if others change the way I have arranged things.    0 1 2 3 4
10. I feel I have to repeat certain numbers.    0 1 2 3 4
11. I sometimes have to wash or clean myself simply because I feel contaminated.    0 1 2 3 4
12. I am upset by unpleasant thoughts that come into my mind against my will.    0 1 2 3 4
13. I avoid throwing things away because I am afraid I might need them later.    0 1 2 3 4
14. I repeatedly check gas and water taps and light switches after turning them off.    0 1 2 3 4
15. I need things to be arranged in a particular order.    0 1 2 3 4
16. I feel that there are good and bad numbers.    0 1 2 3 4
17. I wash my hands more often and longer than necessary.    0 1 2 3 4
18. I frequently get nasty thoughts and have difficulty in getting rid of them.    0 1 2 3 4

## Yale–Brown Obsessive Compulsive Scale (Y-BOCS)

"I am now going to ask several questions about your obsessive thoughts." [Make specific reference to the patient's target obsessions.]

1. **TIME OCCUPIED BY OBSESSIVE THOUGHTS**

Q: How much of your time is occupied by obsessive thoughts? [When obsessions occur as brief, intermittent intrusions, it may be difficult to assess time occupied by them in terms of total hours. In such cases, estimate time by determining how frequently they occur. Consider both the number of times the intrusions occur and how many hours of the day are affected. Ask:] How frequently do the obsessive thoughts occur? [Be sure to exclude ruminations and preoccupations which, unlike obsessions, are ego-syntonic and rational (but exaggerated).]

0 = None.
1 = Mild, less than 1 hr/day or occasional intrusion.
2 = Moderate, 1 to 3 hrs/day or frequent intrusion.
3 = Severe, greater than 3 and up to 8 hrs/day or very frequent intrusion.
4 = Extreme, greater than 8 hrs/day or near constant intrusion.

1b. **OBSESSION-FREE INTERVAL** (not included in total score)

Q: On the average, what is the longest number of consecutive waking hours per day that you are completely free of obsessive thoughts? [If necessary, ask:] What is the longest block of time in which obsessive thoughts are absent?

0 = No symptoms.
1 = Long symptom-free interval, more than 8 consecutive hours/day symptom-free.
2 = Moderately long symptom-free interval, more than 3 and up to 8 consecutive hours/day symptom-free.
3 = Short symptom-free interval, from 1 to 3 consecutive hours/day symptom-free.
4 = Extremely short symptom-free interval, less than 1 consecutive hour/day symptom-free.

2. **INTERFERENCE DUE TO OBSESSIVE THOUGHTS**

Q: How much do your obsessive thoughts interfere with your social or work (or role) functioning? Is there anything that you don't do because of them? [If currently not working determine how much performance would be affected if patient were employed.]

0 = None.
1 = Mild, slight interference with social or occupational activities, but overall performance not impaired.
2 = Moderate, definite interference with social or occupational performance, but still manageable.
3 = Severe, causes substantial impairment in social or occupational performance.
4 = Extreme, incapacitating.

3. **DISTRESS ASSOCIATED WITH OBSESSIVE THOUGHTS**

Q: How much distress do your obsessive thoughts cause you? [In most cases, distress is equated with anxiety; however, patients may report that their obsessions are "disturbing" but deny "anxiety." Only rate anxiety that seems triggered by obsessions, not generalized anxiety or anxiety associated with other conditions.]

0 = None
1 = Mild, not too disturbing
2 = Moderate, disturbing, but still manageable
3 = Severe, very disturbing
4 = Extreme, near constant and disabling distress

4. **RESISTANCE AGAINST OBSESSIONS**

Q: How much of an effort do you make to resist the obsessive thoughts? How often do you try to disregard or turn your attention away from these thoughts as they enter your mind? [Only rate effort made to resist, not success or failure in actually controlling the obsessions. How much the patient resists the obsessions may or may not correlate with his/her ability to control them. Note that this item does not directly measure the severity of the intrusive thoughts; rather it rates a manifestation of health, i.e., the effort the patient makes to counteract the obsessions by means other than avoidance or the performance of compulsions. Thus, the more the patient tries to resist, the less impaired is this aspect of his/her functioning. There are "active" and "passive" forms of resistance. Patients in behavioral therapy may be encouraged to counteract their obsessive symptoms by not struggling against them (e.g., "just let the thoughts come"; passive opposition) or by intentionally bringing on the disturbing thoughts. For the purposes of this item, consider use of these behavioral techniques as_forms of resistance. If the obsessions are minimal, the patient may not feel the need to resist them. In such cases, a rating of "0" should be given.]

0 = Makes an effort to always resist, or symptoms so minimal doesn't need to actively resist
1 = Tries to resist most of the time
2 = Makes some effort to resist
3 = Yields to all obsessions without attempting to control them, but does so with some reluctance
4 = Completely and willingly yields to all obsessions

5. **DEGREE OF CONTROL OVER OBSESSIVE THOUGHTS**

Q: How much control do you have over your obsessive thoughts? How successful are you in stopping or diverting your obsessive thinking? Can you dismiss them? [In contrast to the preceding item on resistance, the ability of the patient to control his obsessions is more closely related to the severity of the intrusive thoughts.]

0 = Complete control
1 = Much control, usually able to stop or divert obsessions with some effort and concentration.
2 = Moderate control, sometimes able to stop or divert obsessions.
3 = Little control, rarely successful in stopping or dismissing obsessions, can only divert attention with difficulty.
4 = No control, experienced as completely involuntary, rarely able to even momentarily alter obsessive thinking.

"The next several questions are about your compulsive behaviors." [Make specific reference to the patient's target compulsions.]

6. **TIME SPENT PERFORMING COMPULSIVE BEHAVIORS**

Q: How much time do you spend performing compulsive behaviors? [When rituals involving activities of daily living are chiefly present, ask:] How much longer than most

but not all cases, performing compulsions reduces anxiety. If, in the judgment of the interviewer, anxiety is actually reduced by preventing compulsions in the manner described above, then ask:] How anxious do you get while performing compulsions until you are satisfied they are completed?

0 = None

1 = Mild only slightly anxious if compulsions prevented, or only slight anxiety during performance of compulsions

2 = Moderate, reports that anxiety would mount but remain manageable if compulsions prevented, or that anxiety increases but remains manageable during performance of compulsions

3 = Severe, prominent and very disturbing increase in anxiety if compulsions interrupted, or prominent and very disturbing increase in anxiety during performance of compulsions

4 = Extreme, incapacitating anxiety from any intervention aimed at modifying activity, or incapacitating anxiety develops during performance of compulsions

9. RESISTANCE AGAINST COMPULSIONS
Q: How much of an effort do you make to resist the compulsions? [Only rate effort made to resist, not success or failure in actually controlling the compulsions. How much the patient resists the compulsions may or may not correlate with his ability to control them. Note that this item does not directly measure the severity of the compulsions; rather it rates a manifestation of health, i.e., the effort the patient makes to counteract the compulsions. Thus, the more the patient tries to resist, the less impaired is this aspect of his functioning. If the compulsions are minimal, the patient may not feel the need to resist them. In such cases, a rating of "0" should be given.]

0 = Makes an effort to always resist, or symptoms so minimal doesn't need to actively resist

1 = Tries to resist most of the time

2 = Makes some effort to resist

3 = Yields to almost all compulsions without attempting to control them, but does so with some reluctance

4 = Completely and willingly yields to all compulsions

10. DEGREE OF CONTROL OVER COMPULSIVE BEHAVIOR
Q: How strong is the drive to perform the compulsive behavior? [Pause] How much control do you have over the compulsions? [In contrast to the preceding item on resistance, the ability of the patient to control his compulsions is more closely related to the severity of the compulsions.]

0 = Complete control

1 = Much control, experiences pressure to perform the behavior but usually able to exercise voluntary control over it

2 = Moderate control, strong pressure to perform behavior, can control it only with difficulty

3 = Little control, very strong drive to perform behavior, must be carried to completion can only delay with difficulty

4 = No control, drive to perform behavior experienced as completely involuntary and overpowering, rarely able to even momentarily delay activity

people does it take to complete routine activities because of your rituals? [When compulsions occur as brief, intermittent behaviors, it may difficult to assess time spent performing them in terms of total hours. In such cases, estimate time by determining how frequently they are performed. Consider both the number of times compulsions are performed and how many hours of the day are affected. Count separate occurrences of compulsive behaviors, not number of repetitions; e.g., a patient who goes into the bathroom 20 different times a day to wash his hands 5 times very quickly, performs compulsions 20 times a day, not 5 or 5 × 20 = 100. Ask:] How frequently do you perform compulsions? [In most cases compulsions are observable behaviors (e.g., hand washing), but some compulsions are covert (e.g., silent checking).]

0 = None

1 = Mild (spends less than 1 hr/day performing compulsions), or occasional performance of compulsive behaviors.

2 = Moderate (spends from 1 to 3 hrs/day performing compulsions), or frequent performance of compulsive behaviors.

3 = Severe (spends more than 3 and up to 8 hrs/day performing compulsions), or very frequent performance of compulsive behaviors.

4 = Extreme (spends more than 8 hrs/day performing compulsions), or near constant performance of compulsive behaviors (too numerous to count).

6b. COMPULSION-FREE INTERVAL (not included in total score)
Q: On the average, what is the longest number of consecutive waking hours per day that you are completely free of compulsive behavior? [If necessary, ask:] What is the longest block of time in which compulsions are absent?

0 = No symptoms.

1 = Long symptom-free interval, more than 8 consecutive hours/day symptom-free.

2 = Moderately long symptom-free interval, more than 3 and up to 8 consecutive hours/day symptom-free.

3 = Short symptom-free interval, from 1 to 3 consecutive hours/day symptom-free.

4 = Extremely short symptom-free interval, less than 1 consecutive hour/day symptom-free.

7. INTERFERENCE DUE TO COMPULSIVE BEHAVIORS
Q: How much do your compulsive behaviors interfere with your social or work (or role) functioning? Is there anything that you don't do because of the compulsions? [If currently not working determine how much performance would be affected if patient were employed.]

0 = None

1 = Mild, slight interference with social or occupational activities, but overall performance not impaired

2 = Moderate, definite interference with social or occupational performance, but still manageable

3 = Severe, causes substantial impairment in social or occupational performance

4 = Extreme, incapacitating

8. DISTRESS ASSOCIATED WITH COMPULSIVE BEHAVIOR
Q: How would you feel if prevented from performing your compulsion(s)? [Pause] How anxious would you become? [Rate degree of distress patient would experience if performance of the compulsion were suddenly interrupted without reassurance offered. In most,

Appendix B

EMA Items

[EMA-OCD (items based on OCI-R)]
[Incidental EMA items]

EMA-OCD (items based on OCI-R)

In the last four hours, indicate if you engaged in each of the following:

1. I saved or collected something I don't need.

NO     Yes

How many times?     ↵

How long did you spend doing this?     ↵

2. I checked something I didn't need to.

NO     Yes

How many times?     ↵

How long did you spend doing this?     ↵

3. I got upset if objects were not arranged properly.

NO     Yes

How many times?     ↵

How long did you spend doing this?     ↵

4. I found myself repeatedly counting objects.

NO     Yes

How many times did you do this?     ↵

How long did you spend doing this?     ↵

5. I found myself not wanting to touch an object touched by someone else.

NO     Yes

How many times did this happen?     ↵

How long did you spend doing this?     ↵

6.  I found it difficult to control my own thoughts.    NO    Yes

How many times did this happen?    ↵

How long did you spend doing this?    ↵

7.  I collected things I don't need.    NO    Yes

How many times did you do this?    ↵

How long did you spend doing this?    ↵

8.  I repeatedly checked doors, windows, drawers, etc.    NO    Yes

How many times did you do this?    ↵

How long did you spend doing this?    ↵

9.  I got upset when someone changed the way I arranged things.    NO    Yes

How many times did this happen?    ↵

How long did you spend doing this?    ↵

10. I repeated certain numbers.    NO    Yes

How many times did you do this?    ↵

How long did you spend doing this?    ↵

11. I washed or cleaned myself because I felt contaminated.    NO    Yes

How many times did you do this?    ↵

How long did    ↵

you spend
doing this?

12. Unpleasant thoughts came into my mind against my will.  **NO**  **Yes**

How many
times did this  ↵
happen?

How long did
you spend  ↵
doing this?

13. I avoided throwing something away because I was afraid I might need it later.  **NO**  **Yes**

How many
times did this  ↵
happen?

How long did
you spend  ↵
doing this?

14. I checked stove burners, water taps, and/or light switches after turning them off.  **NO**  **Yes**

How many
times did you  ↵
do this?

How long did
you spend  ↵
doing this?

15. I arranged things in a particular order.  **NO**  **Yes**

How many
times did you  ↵
do this?

How long did
you spend  ↵
doing this?

16. I thought about good and bad numbers.  **NO**  **Yes**

How many
times did you  ↵
do this?

How long did
you spend  ↵
doing this?

17. I washed my hands longer than necessary.                    NO          Yes

How many
times did you    ↵
do this?

How long did
you spend         ↵
doing this?

18. I had troubling thoughts that were difficult to get rid of.          NO          Yes

How many
times did this    ↵
happen?

How long did
you spend          ↵
doing this?

**Supplemental EMA Items**

Environmental variables
Where are you? [Drop down menu]

- work
- school
- home
- friend's home
- relative's home
- public place
- other

In the last 4 hours, what percent of time were you with someone? [Drop down menu]

- 100%
- 80-99%
- 60-79%
- 40-59%
- 20-39%
- 1-19%
- 0% (alone the whole time)

Supplemental variables
How many hours did you sleep last night? [only at the first assessment of the day]

Rate your average mood over the last 4 hours

       Positive -------------------------------------------------------------------------- Negative

Rate your average stress level over the last 4 hours

       Not      -------------------------------------------------------------------   Very
       Stressed                                                              Stressed

Rate your average anxiety level over the last 4 hours

       Not      -------------------------------------------------------------------   Very
       Anxious                                                              Anxious

Rate your average loneliness level over the last 4 hours

       Not      -------------------------------------------------------------------   Very
       Lonely                                                              Lonely

In the past 4 hours, did you have a fight/disagreement with someone?
       Yes    No
If yes: How upsetting was it for you?

       Not at all-----------------------------------------------------------------------Very
       Upsetting                                                                 Upsetting

Appendix C

Time 3 Materials/ Assessments

[Correlation Tutorial]
[Correlation Scales]

# Correlation Tutorial

## OCD Assessment

We will soon ask you to make some ratings about your behavior during the previous week. Before we do this, we need to briefly discuss a concept called *correlation.*

# What is a correlation?

- A *correlation* is a number that describes how two things are related to each other.
- There are several types of possible relations. Let's talk about a few…

# In one type of relation…

- As one thing goes up or increases, so does the other.
- For example…
  - When someone runs faster ↑, their heartbeat speeds up ↑
    - So we say: "Running speed is *positively correlated* with heartbeat."
  - The happier somebody feels ↑, the more they smile ↑
    - So we say: "Happiness is *positively correlated* with smiling."
  - The better someone is at math ↑, the faster they can balance their checkbook ↑
    - So we say: "Math ability is *positively correlated* with speed of balancing a checkbook."
- Each of these is an example of a **positive correlation**

# In another type of relation…

- As one thing goes up, the other goes down.
- For example…
  - As smoking increases ↑, life-expectancy decreases ↓
    - So we say: "Smoking is *negatively correlated* with life-expectancy."
  - The more TV someone watches ↑, the lower their score on a class exam ↓
    - So we say: "Watching TV is *negatively correlated* with exam grades."
- Each of these is an example of a **negative correlation**

# In the final type of relation…

- There is actually <u>no relation</u> at all between two things.
- Knowing the value of one thing tells us nothing about the value of the other
- For example…
  - If shoe size goes up ↑, we learn nothing new about how intelligent someone may be
    - So we say: "Shoe size is *not correlated* with intelligence"
  - The color of a car tells us nothing about how fast it goes
    - So we say: "Car color is *not correlated* with car speed."
- Each of these is an example of **<u>no correlation</u>** between two things

# Rating a relationship

- We can show the strength of a relationship by rating it from -100 to +100.

-100  -90  -80  -70  -60  -50  -40  -30  -20  -10  0  +10  +20  +30  +40  +50  +60  +70  +80  +90  +100

Strongest possible
**negative relation**
between two
things

No relationship
between two
Things
(**no correlation**)

Strongest possible
**positive relation**
between two
Things

(**negative correlation**):
*As one thing goes up* ↑
*The other goes down* ↓

(**positive correlation**):
*As one thing goes up* ↑,
*so does the other* ↑
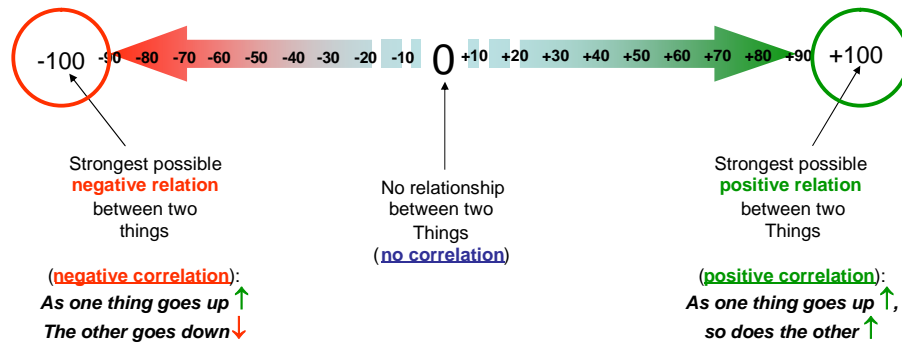
# A correlation of "0"

- A rating of "0" would mean that there is no relationship between two things: they are not correlated
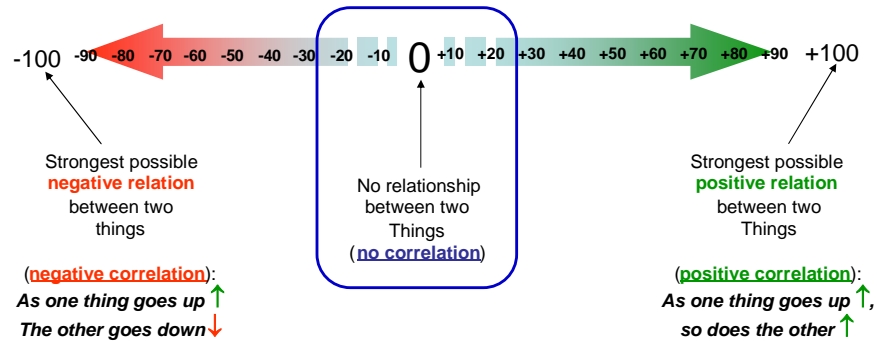
# Higher number = Stronger relation

- As the number increases (either positively or negatively), the relation is stronger and more predictable. As the number decreases (and gets closer to zero), the relation is weaker.

-100    -90  -80  -70  -60  -50  -40  -30  -20  -10    0    +10  +20  +30  +40  +50  +60  +70  +80  +90    +100

Strongest possible
**negative relation**
between two
things

(**negative correlation**)
*As one thing goes up* ↑
*The other goes down* ↓

No relationship
between two
Things
(**no correlation**)

Strongest possible
**positive relation**
between two
Things

(**positive correlation**):
*As one thing goes up* ↑,
*so does the other* ↑
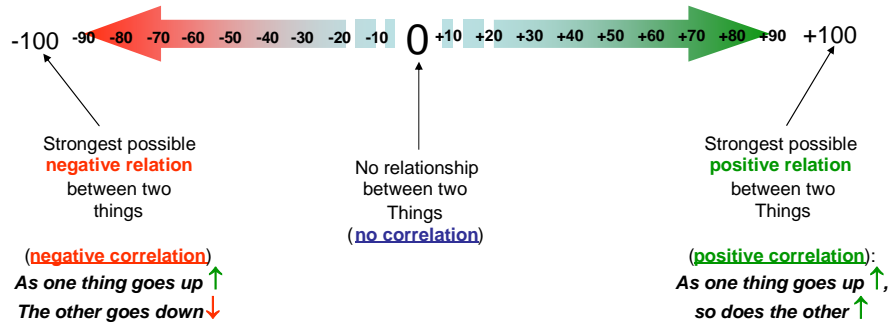
# Notice…

- …Negative relations are on the left (red)
- …Positive relations are on the right (green)
- …No relationship is in the middle (blue)

-100  -90 -80 -70 -60 -50 -40 -30 -20 -10  0  +10 +20 +30 +40 +50 +60 +70 +80 +90  +100

*As one thing goes up* ↑
*The other goes down* ↓

*As one thing goes up* ↑,
*so does the other* ↑

No relationship
between two
things
(**no correlation**)
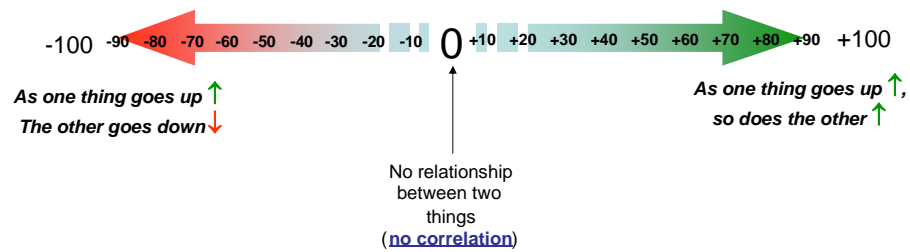
# What the numbers mean

- The **bigger** the number (positive *or* negative), the **stronger** and more predictable the relationship
  - So, a +90 relation is stronger, and more predictable, than +30
    - It's a stronger *positive* relationship
  - Similarly, a -90 relation is stronger, and more predictable, than -30
    - It's a stronger *negative* relationship

-100   -90  -80  -70  -60  -50  -40  -30  -20  -10   0   +10  +20  +30  +40  +50  +60  +70  +80  +90   +100

Weaker
Relationships
(closer to zero)

Stronger (more predictable)
relationships are
further from zero and
have bigger numbers
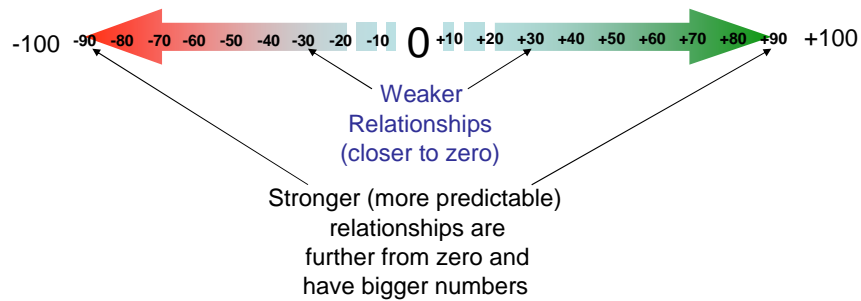
# Let's try some examples

- We might think that the relationship between **height** and **weight** is positive and very predictable (taller people almost always weigh more)
- So we might estimate its relationship to be high and positive: +80

-100  -90  -80  -70  -60  -50  -40  -30  -20  -10  0  +10  +20  +30  +40  +50  +60  +70  +80  +90  +100
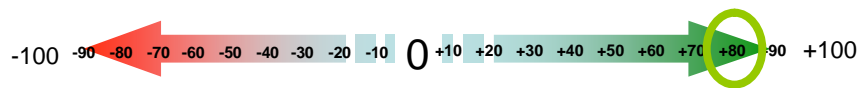
# Here's another example

- Conversely, we might think that the relationship between **TV watching** and **class grades** is negative and very predictable (as someone watches more TV, there grades go down)

- So we might estimate its relationship to be high and negative: - 90

-100  -90  -80  -70  -60  -50  -40  -30  -20  -10  0  +10  +20  +30  +40  +50  +60  +70  +80  +90  +100
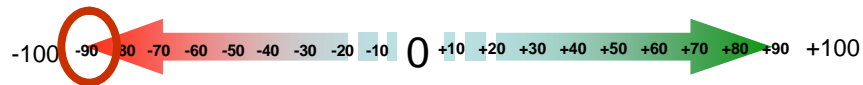
# Examples

- What if there's no relationship?
- Remember our earlier example of a car's speed and its color?
- Since there is no relationship between a car's speed and its color (knowing a car's color tells us nothing about how fast it can go), we'd circle zero.

-100  -90  -80  -70  -60  -50  -40  -30  -20  -10  0  +10  +20  +30  +40  +50  +60  +70  +80  +90  +100

# But what about weaker relationships that are less predictable?

- When a relationship is weaker, the number associated with it should be smaller since the relationship is not as predictable.
- For example, people who work hard sometimes get praised for the efforts, sometimes not. We might think of this relationship as positive but weak: +20

-100   -90  -80  -70  -60  -50  -40  -30  -20  -10   0   +10  +20  +30  +40  +50  +60  +70  +80  +90   +100
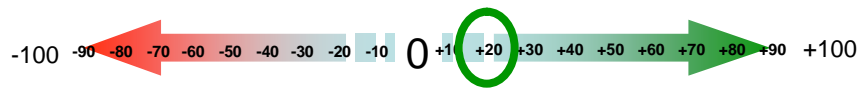
# A weak, negative relationship

- Similarly, some relationships are negative *and* weak.
- For example, how much a person exercises may be weakly related to how stressed the person feels.
- As exercise goes up, stress goes down
- However, since exercise by itself may not remove all of a person's stress, we might rate it: -30

-100  -90 -80 -70 -60 -50 -4 -30 20 -10  0 +10 +20 +30 +40 +50 +60 +70 +80 +90  +100
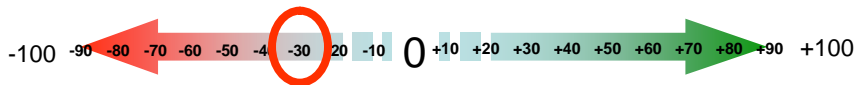
Now we'll practice some of these ideas.

# What is the relation between the length of someone's foot and their shoe size?
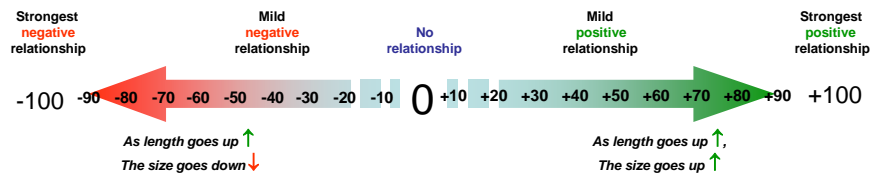
| Strongest negative relationship | Mild negative relationship | No relationship | Mild positive relationship | Strongest positive relationship |
|---|---|---|---|---|

-100  -90  -80  -70  -60  -50  -40  -30  -20  -10  0  +10  +20  +30  +40  +50  +60  +70  +80  +90  +100

*As length goes up* ↑
*The size goes down* ↓

*As length goes up* ↑,
*The size goes up* ↑

**Click one of the buttons below:**

▶ **Between -51 and -100**

▶ **Between -1 and -50**

▶ **0**

▶ **Between +1 and +50**

▶ **Between +51 and +100**

# Correct!

- That is correct! As the length of someone's foot increases, their shoe size increases. This is a strong positive relation, so we circle a number on the far right of this scale.
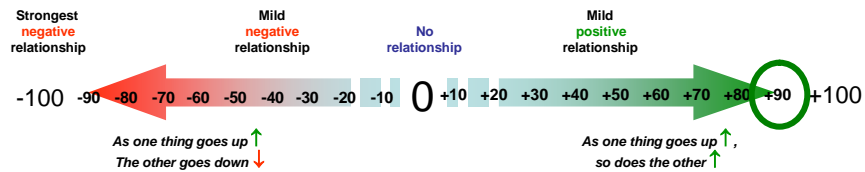


Press here to move to the next question.

# Incorrect

- Please try again.

Click here to try again.

# What is the relation between how healthy you eat and how often you go to the doctor?
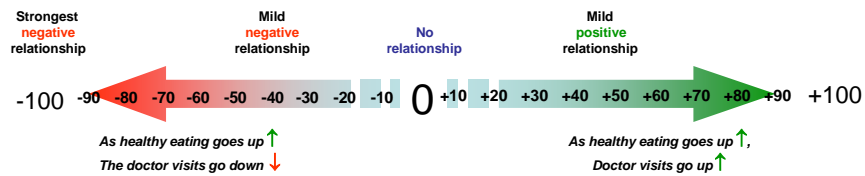
**Strongest negative relationship**   **Mild negative relationship**   **No relationship**   **Mild positive relationship**

-100  -90 -80 -70 -60 -50 -40 -30 -20 -10  0  +10 +20 +30 +40 +50 +60 +70 +80 +90  +100

*As healthy eating goes up ↑*
*The doctor visits go down ↓*

*As healthy eating goes up ↑,*
*Doctor visits go up ↑*

**Click one of the buttons below:**

▶ **Between -51 and -100**

▶ **Between -1 and -50**

▶ **0**

▶ **Between +1 and +50**

▶ **Between +51 and +100**

# Correct!

- That is correct! Healthy eating usually is associated with fewer doctor visits. It is a mild negative relationship because eating healthy is *not always* associated with fewer visits to the doctor.
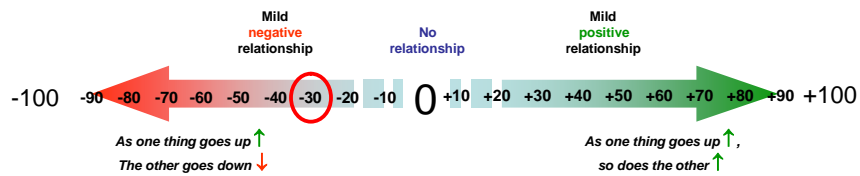


Press here to move to the next question.

# Incorrect

- Please try again.

 Click here to try again.

# What is the relation between the number of children in your family and length of a football field?
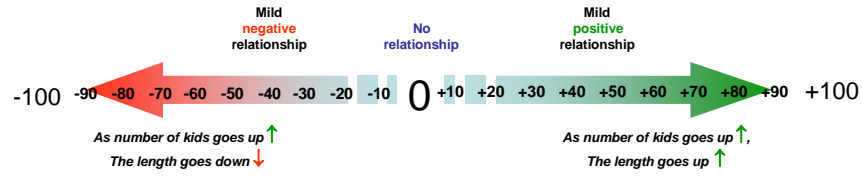
**Mild negative relationship**

**No relationship**

**Mild positive relationship**

-100  -90 -80 -70 -60 -50 -40 -30 -20 -10  0  +10 +20 +30 +40 +50 +60 +70 +80 +90  +100

*As number of kids goes up ↑*
*The length goes down ↓*

*As number of kids goes up ↑,*
*The length goes up ↑*

**Click one of the buttons below:**

▶ **Between -51 and -100**

▶ **Between -1 and -50**

▶ **0**

▶ **Between +1 and +50**

▶ **Between +51 and +100**

# Correct!

- That is correct! There is no relation between the number of children in your family and the length of a football field. Since there is no relationship, we circle the number 0.

| | | |
|---|---|---|
| **Mild** <br> **negative** <br> relationship | **No** <br> **relationship** | **Mild** <br> **positive** <br> relationship |

-100  -90  -80  -70  -60  -50  -40  -30  -20  -10  〇  +10  +20  +30  +40  +50  +60  +70  +80  +90  +100

*As one thing goes up* ↑ <br> *The other goes down* ↓

*As one thing goes up* ↑ , <br> *so does the other* ↑
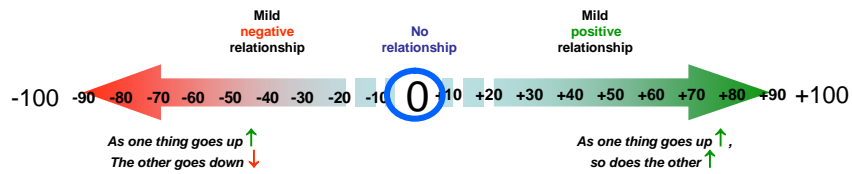
▶ Press here to move to finish.

# Incorrect

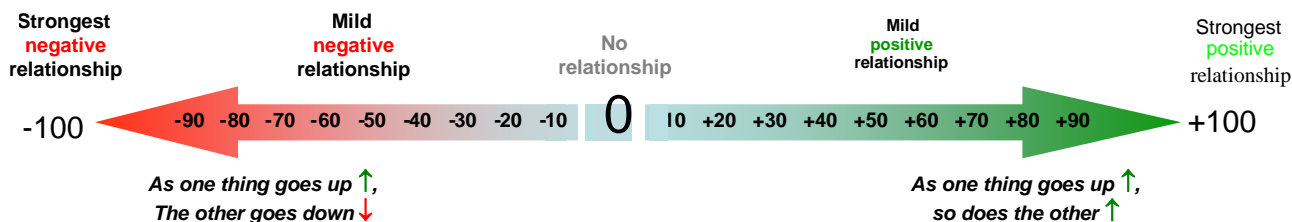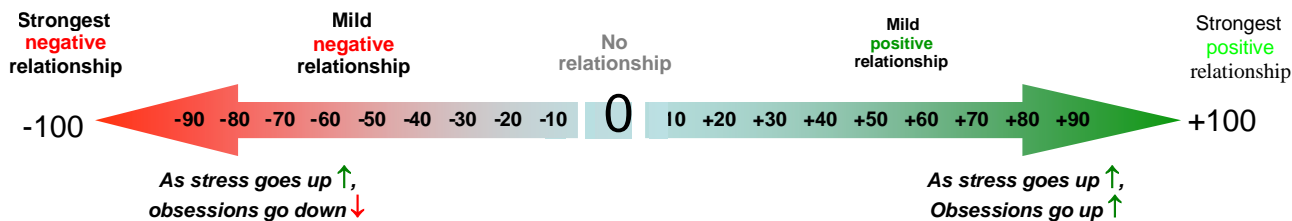- Please try again.

Click here to try again.

# GREAT!!!

- Now that you understand this concept, we have some questions about different thoughts and behaviors you experienced in the last week…

Time 3 Assessments

Participant estimates of symptom covariance will be assessed using the following scale:

**Strongest** **negative** **relationship**  **Mild** **negative** **relationship**  **No** **relationship**  **Mild** **positive** **relationship**  **Strongest** **positive** relationship

-100  -90  -80  -70  -60  -50  -40  -30  -20  -10  0  10  +20  +30  +40  +50  +60  +70  +80  +90  +100

*As one thing goes up ↑,*  *As one thing goes up ↑,*
*The other goes down↓*  *so does the other ↑*

---

Example:
Please circle a number that best indicates the relationship between Stress and Obsessions

**Strongest** **negative** **relationship**  **Mild** **negative** **relationship**  **No** **relationship**  **Mild** **positive** **relationship**  **Strongest** **positive** relationship

-100  -90  -80  -70  -60  -50  -40  -30  -20  -10  0  10  +20  +30  +40  +50  +60  +70  +80  +90  +100

*As stress goes up ↑,*  *As stress goes up ↑,*
*obsessions go down↓*  *Obsessions go up ↑*

This was repeated for obsessing and the two OCI-R subscales on which participants scored highest (i.e., washing, hoarding, ordering, checking, or neutralizing) and the supplemental items (sleep, being alone, stress, positive mood, negative mood, loneliness, anxiety, & distressing fights)

1. During the last week, did you ever not tell the truth when answering the questions?

2. How normal was the week for you?

3. What percentage of the beeps did you respond to?

4. Did recording with the hand-held computer change your symptoms?
   a) Not at all
   b) A little
   c) A lot
   d) My symptoms are gone because of self-monitoring

5. How confident are you in the accuracy of your memory?

6. How confident are you in the accuracy of the information you entered in the handheld computer?

7. How confident are you in the estimates you just provided?

8. Did you think that we were going to ask you questions about your memory?
    a) If yes: When did you suspect this? Why?

9. Would you participate in a study like this again or recommend participation to a friend?

Appendix D

Glossary

*Definitions of technical terms*

| Term | Definition |
| --- | --- |
| Autobiographical memory | Memories of one's own past behavior |
| Bias | Systematic error |
| Episodic memory | Memories of personal events dated in the past. Can be contrasted from semantic memory (i.e., memory of facts) and procedural memory (i.e., actions that are relatively automatic and not open to introspection.) |
| Explicit memory | Memories of events characterized by conscious recall. |
| Implicit memory | Memories characterized by the lack of consciousness in the act of recalling. It is often "remembered" as part of an action and demonstrated by improvement in a procedural task. |
| Inaccurate | Random error |
| Overgenerality | Tendency to recall categories of events as opposed to specific instances |
| Prospective Recall | The act of remembering to engage in future activities |
| Retrospective Recall | The act of remembering memories of past events |
| Recall task | Producing an item from memory without a clue |
| Recognition task | Identifying target items from distracter items |
| Schema | Abstract representations or prototypical scripts of event types, from which accounts of particular instances are derived when recall is solicited |

Script-like behavior          Behaviors that follow predictable, if slightly varied

patterns such as birthdays and baseball games.

Appendix E
Human Subjects Review Committee Approval Forms