Author's personal copy

Journal of Obsessive-Compulsive and Related Disorders 1 (2012) 331–335

Contents lists available at SciVerse ScienceDirect

Journal of Obsessive-Compulsive and Related Disorders

journal homepage: www.elsevier.com/locate/jocrd

Good news for allegedly bad studies. Assessment of psychometric properties may help to elucidate deception in online studies on OCD

Steffen Moritz [a,*], Niels Van Quaquebeke [b], Marit Hauschildt [a], Lena Jelinek [a], Sascha Gönner [c]

[a] University Medical Center Hamburg-Eppendorf, Department of Psychiatry and Psychotherapy, D-20246 Hamburg, Germany
[b] Kuehne Logistics University, Department of Management and Economics, Hamburg, Germany
[c] Study Centre for Behavioural Medicine and Psychotherapy, Stuttgart, Germany

ARTICLE INFO

Article history:
Received 7 April 2012
Received in revised form
14 June 2012
Accepted 4 July 2012
Available online 24 July 2012

Keywords:
OCD
Online studies
Survey
Internet
Deception
Simulation

ABSTRACT

Online surveys are gaining increasing momentum in clinical research. Ease of recruitment and low cost are two of the biggest advantages of Internet studies. There are, however, concerns about their reliability and validity.

The present study compared the psychometric properties of self-report instruments measuring obsessive-compulsive disorder (OCD) across three samples: (1) participants with a confirmed diagnosis of OCD ($n=66$), (2) participants with a probable diagnosis of OCD ($n=86$) and (3) clinical experts on OCD and students who were asked to pretend to have OCD ($n=121$). Psychometric indices of the Yale-Brown Obsessive Compulsive Score (Y-BOCS) and the Obsessive Compulsive Inventory (OCI-R) served as indicators for reliability and validity.

Both patient samples revealed good retest reliability scores and good correlations between Y-BOCS and OCI-R scores. In contrast, the expert group showed poor retest reliabilities and mixed results for the intercorrelations between OCI-R and Y-BOCS scores. Simulators display a marked tendency to over-report symptoms on the OCI-R.

Good psychometric properties of online studies may serve as a proxy for the validity of diagnoses.
© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Intervention trials conducted via the Internet are a cost-effective tool that is increasingly adopted in clinical psychological and psychiatric studies (for reviews see for example Cuijpers et al., 2011; Moritz, Timpano, Wittekind, & Knaevelsrud, in press). This line of research is particularly useful at the early stages of a project, at which external funding is difficult to obtain (i.e., proof-of-concept). Moreover, anonymous Internet surveys are especially beneficial in reaching and providing treatment for people with psychological problems who are not yet willing or able to approach the psychiatric-psychological help system (for reviews see Moritz, Wittekind, Hauschildt, & Timpano, 2011). For several psychological disorders, such as obsessive-compulsive disorder (OCD) and depression, those who do not actively seek face-to-face (FTF) treatment even represent the majority (Kohn, Saxena, Levav, & Saraceno, 2004). Internet studies are thus an important complement to (albeit not a substitute for) traditional clinical studies. Moreover, emerging evidence suggests that the clinical characteristics of those who seek FTF treatment differ from those who do not (e.g., Besiroglu, Cilli, & Askin, 2004; Brett, Johns, Peters, & McGuire, 2009). Hence, findings obtained in clinical studies cannot be easily extrapolated to the entire population and Internet studies are an important means to fill this empirical gap.

At the same time, anonymous Internet surveys often raise concerns about reliability and validity (for a summary see Hancock, 2007). Without a formal diagnosis determined by an expert it cannot be excluded that some individuals fake having a disorder (e.g., in order to sabotage a project or receive (monetary) compensation for participation). While this possibility is considered to be low (Buchanan et al., 2005; Whitehead, 2007), there is no solid data on the extent of deception and its impact on the results. Lie scales and control questions are means to reduce such "noise" but may not entirely rule out this objection (see discussion).

While not speaking about clinical populations per se, a first review (Gosling, Vazire, Srivastava, & John, 2004) comparing personality data collected with large Internet samples against a set of 510 traditional studies (published in the *Journal of Personality and Social Psychology*) on important statistical quality indicators such as reliability, non-response, discriminant intercorrelations, and nonsense data entries comes to the conclusion that generally web-based survey data can be trusted. Indeed, Gosling et al. were able to show that Internet samples are as

* Corresponding author. Tel.: +49 40 7410 56565; fax: +49 40 7410 57566.
E-mail address: moritz@uke.uni-hamburg.de (S. Moritz).

2211-3649/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jocrd.2012.07.001

Author's personal copy

diverse as traditional samples and not composed of maladjusted, socially isolated, or depressed people. Further they found online participants not to be in any way less motivated to take studies seriously and respond meaningfully compared with regular testing methods. Also, they were able to show that the anonymity provided by web questionnaires does not compromise data quality when measures such as full disclosure of potential feedback (or no feedback at all) are employed. Hence in sum, they arrive at a rather positive assessment of online research: "Our analyses also suggest that the data provided by Internet methods are of at least as good quality as those provided by traditional paper-and-pencil methods. In short, the data collected from Internet methods are not as flawed as is commonly believed." (Gosling et al., 2004, p. 102). Similarly, even Hancock (2007), while critically noting the possibilities of deception in the online sphere, acknowledges in his review that deception may occur at least to the same extent in FTF communication. Still, these inferences are based on studies with non-clinical samples and data on clinical subjects are still lacking.

The present study investigated robustness against deception in clinical Internet studies. To meet this purpose, we chose OCD, a disorder characterized by repetitive and bothersome thoughts (i.e. obsessions such as concerns about dirt) and neutralizing behavior (i.e. compulsions such as washing). OCD patients are of particular interest for online studies because of their reluctance to seek FTF treatment (Kohn et al., 2004), partly owing to OCD-specific concerns (e.g., fear of being regarded as criminal because of aggressive thoughts). In particular, we examined whether people who actively simulate having OCD can be distinguished from true patients with respect to the psychometric indices. For example, it has been put forward that unreliable and random responses should manifest in diminished internal reliability scores (Gosling et al., 2004). We think it is high time for a study investigating the reliability of online research in OCD given the growing number of OCD studies conducted over the Internet (e.g., Coles, Johnson, & Schubert, 2011; Marques et al., 2010; Mataix-Cols & Marks, 2006; Wootton, Titov, Dear, Spence, & Kemp, 2011).

For our study, we compared three samples recruited via the Internet: (1) patients with a verified diagnosis of OCD, (2) subjects with a probable diagnosis of OCD, and (3) academics and students with varying grades of expertise on OCD. The latter were encouraged to fill out OCD questionnaires, which have been tested for use in online trials (e.g., Coles, Cook, & Blake, 2007; Moritz et al., 2011), as if they had OCD. Each study involved a baseline assessment and a four-week post survey.

We expected that the psychometric properties in samples 1 and 2 would be comparable to results obtained in clinical studies. In contrast, we predicted that sample 3 (experts simulating OCD) who were forewarned that we aimed to assess the reliability of their responses would still show low test-retest reliability, as simulators may not be able to recollect at long retest intervals which responses they gave in the baseline assessment.

2. Methods

2.1. Participants

Three different samples were recruited in German-speaking countries. Each sample was assessed twice with a four-week interval in between. Baseline assessments and post-assessment instruments are described below. Informed consent was obtained online from the participants in accordance with the local department of data security and the local Ethics committee.

Sample 1 (OCD patients with verified diagnoses) consisted of 66 patients who took part in an ongoing trial on the efficacy of metacognitive training in OCD (myMCT, www.uke.de/mymct Moritz & Hauschildt, 2011), a program aimed at attenuating the cognitive biases and dysfunctional beliefs putatively involved in the formation and maintenance of OCD (e.g., inflated responsibility, perfectionism). Patients were recruited via online and self-help forums. Additional subjects were recruited with the aid of the German association for OCD (DGZ). Individuals were randomly allocated to either the experimental condition (myMCT) or a control group who received psychoeducative information (the latter group will be analyzed separately, see below). Inclusion criteria were age between 17 and 70 and a primary diagnosis of OCD. Core exclusion criteria were psychotic or bipolar disorders. In addition to the online assessment, each patient had to undergo a telephone interview at baseline and four weeks later. The interviewer was blind to treatment status. The interviewer verified diagnostic status of OCD according to the Mini Neuropsychiatric Interview (MINI; Sheehan et al., 1998) and also administered the Yale-Brown Obsessive Compulsive Score exert-rating (Y-BOCS; Goodman et al., 1989). Among other instruments, the online survey requested the Y-BOCS self-rating version (see below).

Sample 2 (OCD patients without verified diagnoses) consisted of 86 patients who took part in a study which dealt with the efficacy of a forerunner version of the myMCT (Moritz & Hauschildt, 2011). The participants were randomly allocated to either the intervention condition or a wait-list control group. Inclusion criteria were a prior diagnosis of OCD according to an expert. However, unlike in sample 1, diagnostic status was not formally verified via interview. The core results of this study have been previously published (Moritz, Jelinek, Hauschildt, & Naber, 2010). Recruitment sources were similar for samples 1 and 2 (see above).

Sample 3 (experts simulating a diagnosis of OCD) consisted of 121 non-clinical individuals with an academic background in the mental health system. These participants had to fill out questionnaires on OCD, having been explicitly instructed to complete the questionnaires as if they had OCD. There were no constraints on whether participants took the role of, for example, a patient with checking or washing compulsions. In addition to the questionnaires filled out by all participants (samples 1–3), participants in sample 3 were asked about their expert status, knowledge about OCD, and source of expertise (e.g., from original articles, daily work, relatives/friends suffering from OCD) at the beginning. Unlike recruitment for samples 1 and 2, we approached specific persons in order to ensure that only persons with sufficient expertise would take part in the survey. The first author contacted medical students at the University Medical Center in Hamburg from three different trimesters who had either attended lessons on psychiatric illnesses before or were currently attending lessons on psychopathology. We also approached psychology students who attended a seminar on OCD held by the first author as well as students from his laboratory. We also contacted experts who either work with and/or conduct research on people with OCD (mainly psychologists and/or psychiatrists). Eighty-one participants were considered to be "distinguished" experts as they were actively engaged in research on OCD, worked with OCD patients and/or had read or written original research articles on OCD. To make the study more conservative, we disclosed the purpose of the study to these participants beforehand.

2.2. Instruments

Scales were administered in German language using authorized translations. The German instruments fully corresponded to the original instruments with respect to item content, instruction and mode of administration. All participants were requested to fill out the Obsessive-Compulsive Inventory-Revised online (OCI-R) (Foa et al., 2002). The OCI-R is a self-report scale, which assesses the frequency and the degree of distress experienced by OCD symptoms across six subscales, each covered by three items. The OCI-R has good psychometric properties (Abramowitz & Deacon, 2006; Foa et al., 2002; Huppert et al., 2007), especially with respect to the total score, and is sensitive to change (Abramowitz, Tolin, & Diefenbach, 2005). Good psychometric properties have been asserted for the German version (Gönner, Leonhart, & Ecker, 2007, 2008). The equivalence of the original paper version and computer/Internet versions has been established in a study by Coles et al. (2007).

All three samples completed the self-report version of the Yale-Brown Obsessive Compulsive Scale (Y-BOCS) online (Baer, Brown-Beasley, Sorce, & Henriques, 1993; Goodman et al., 1989). The Y-BOCS measures the severity of obsessions and compulsions. The total score is computed from the sum of the ten items (obsessive thoughts: items 1–5; compulsions: items 6–10). The self-report version of the scale has shown strong convergent validity with the original interview version (Schaible, Armbrust, & Nutzinger, 2001; Steketee, Frost, & Bogart, 1996). For the sample with verified diagnosis (sample 1), we also conducted the Y-BOCS interview (Goodman et al., 1989).

2.3. Additional questions for experts

At reassessment, we asked the experts (sample 3) how well they felt they had mastered the task (five-point scales ranging from "very well" (=1) to "very badly" (=5)). We also asked if they had encountered problems (yes/no), such as difficulties in recollecting the responses provided at the baseline assessment (yes/no) and/or difficulties due to lack of expertise (yes/no).

Author's personal copy

Table 1
Baseline differences for the three samples. Frequency, means and standard deviations.

Variables	Sample 1 (verified OCD diagnosis)	Sample 2 (unverified OCD diagnosis)	Sample 3 (simulators)	Statistics (post-hoc tests are Bonferroni-adjusted)
Sociodemographic variables				
Sex (male/female)	26/40	28/58	37/84	$\chi^2(1)=1.53$, $p>.4$
Age in years	40.45 (10.94)	34.52 (10.66)	29.78 (7.37)	$F(2,270)=27.80$; $p<.001$; Sample 1 > Sample 2 > Sample 3
School education (high school level, yes vs. no)	41/25	46/40	121/0	$\chi^2(1)=69.45$, $p<.001$; Sample 3 > Samples 1 and 2
Psychopathology				
Y-BOCS total	19.50 (5.79)	19.29 (6.40)	22.64 (5.72)	$F(2,270)=10.08$; $p<.001$; Sample 3 > Samples 1 and 2
OCI-R total	24.21 (11.27)	28.73 (12.58)[a]	40.36 (11.81)	$F(2,270)=46.42$; $p<.001$; Sample 3 > Samples 1 and 2

Notes: OCI-R=Obsessive-Compulsive Inventory-Revised; Y-BOCS=Yale-Brown Obsessive-Compulsive Scale.

[a] The sum scores of the OCI-R total score in sample 2 deviate from the values reported in the original publication, where OCI-R scores were falsely computed (mistakenly, for some items the lowest value was set higher than 0).

2.4. Strategy of data analysis

We planned to calculate several psychometric properties of the scales including internal consistency (Cronbach's alpha), intercorrelations of the different scales at baseline (external and discriminant validity) as well as test-retest reliabilities. An alpha level of .05 (two-tailed) was used for all statistical tests.

3. Results

3.1. Background and psychopathology

Table 1 shows that gender was equally distributed across the three samples. However, owing to the large number of psychology and medical students in sample 3 (simulators), this sample was considerably younger than the two samples consisting of patients. As this sample only contained academics, formal education level was higher in this group.

3.2. Psychometric data

Table 2 shows data for the three samples as well as subsamples. For sample 3 the subsample comprised the "distinguished" experts (i.e. colleagues with a degree in psychology or medicine who are clinicians and/or researchers working with OCD patients). For samples 1 and 2 the subsamples consisted of patients who were in the control group as the psychometric properties are presumably better in a group with rather stable symptomatology than in the treatment group, where some patients may benefit from intervention while others do not. The three samples performed equally well with respect to internal consistency.

The simulators (sample 3) had somewhat higher mean scores than the two other samples on the Y-BOCS total score (Table 1). However, they were still within the expected range for clinical samples. In contrast, the OCI-R the scores of the simulators were much higher than would be expected for an OCD sample. In addition, the retest reliability was far lower in sample 3, particularly for the Y-BOCS (Table 2)[1]. In the distinguished experts subsample this score was even worse (see brackets in Table 2). The intercorrelation of OCI-R and Y-BOCS scores was higher in the OCD samples. The correlation between the Y-BOCS obsessions and compulsions sub-scores was far lower in the patient samples than for the simulators (see discussion).

[1] We only explored the OCI-R total score as the reliability of the subscores can be low even in clinical individuals.

In sample 3, 89% endorsed that they had problems remembering their initial scores. Lack of expertise as a factor for low performance was reported by 12%.

4. Discussion

The study investigated the robustness of Internet trials against deception, a common preconception against online studies in general (Gosling et al., 2004). We were particularly interested in the question whether "liars" perform differently from true patients. As expected, the two Internet surveys on patients with a likely (sample 2) or a verified diagnosis of OCD (sample 1) yielded comparable psychometric results that were compatible with findings from clinical studies (see below). In contrast, sample 3, which involved non-clinical subjects with an academic background in health care who were encouraged to simulate having a diagnosis of OCD, differed markedly from the two other samples on several plausibility indices. Most importantly, the retest reliability was low, particularly for the Y-BOCS scores (with .27). Prior studies on the Y-BOCS have secured retest reliabilities of .61 (OCD sample, 48.5 day interval; Woody, Steketee, & Chambless, 1995), .88 (nonclinical sample, 1 week; Steketee et al., 1996) and .90 (OCD sample, 1 week; Kim, Dysken, & Kuskowski 1990. Most of the subjects in sample 3 (89%) noted that they had problems at retest remembering what they had entered in the baseline assessment. That the distinguished experts achieved even somewhat lower retest scores on the Y-BOCS could be due to several factors. For example, experts likely know more patients with OCD who might have served as models for their responses. At retest, they may thus have confused these sources, that is, they mimicked the psychopathology of another person at retest. Lack of expertise was reported by just 12% of the participants.

On a cross-sectional level, several indicators for deception emerged. Firstly, the simulators achieved substantially higher scores than the other groups on the OCI-R (over-reporting) but not the *Y*-BOCS. The results obtained in our patient samples accord with most prior research, with OCI-R scores ranging from 24 to 29 points, while the simulators achieved scores that were approximately one standard deviation higher than the means obtained in reference studies (Abramowitz & Deacon, 2006; Abramowitz et al., 2005; Foa et al., 2002; Gönner et al., 2007; Huppert et al., 2007; Sica et al., 2009; Sulkowski et al., 2008). Secondly and perhaps more importantly, the correlation between the Y-BOCS obsessions and compulsions subscores was far lower

Author's personal copy

Table 2
Psychometric properties of the three samples. Results for the control groups are set in square brackets (see text).

Variables	Sample 1 (verified OCD diagnosis; square brackets: control subgroup)	Sample 2 (unverified OCD diagnosis; square brackets: control subgroup)	Sample 3 (simulators; square brackets: subgroup of distinguished experts)
Internal consistency Y-BOCS (Cronbach's α)	.81 [.76]	.80 [.78]	.85 [.86]
Internal consistency OCI-R (Cronbach's α)	.86 [.88]	.86 [.86]	.86 [.87]
Retest reliability Y-BOCS	.68 [.94]	.68 [.82]	.27 [.18]
Retest reliability OCI-R	.75 [.88]	.85 [.84]	.56 [.52]
Correlation between Y-BOCS obsessions and compulsions	.15 [.18]	.12 [.07]	.53 [.59]
Correlation between Y-BOCS and OCI-R	.59 [.55]	.67 [.56]	.44 [.42]

Notes: OCI-R = Obsessive-Compulsive Inventory-Revised; Y-BOCS = Yale-Brown Obsessive-Compulsive Scale. Correlations were significantly different between sample 3 versus samples 2 and 1 for the two retest scores as well as the correlation between Y-BOCS obsessions and compulsions. Sample 3 only differed from sample 2 on the correlation between Y-BOCS and OCI-R (all $p < .05$).

in the patient samples than for simulators. This is in agreement with a number of factor analytic studies indicating that obsessions and compulsions are only weakly correlated and that the Y-BOCS is best represented by two (obsessions versus compulsions; Boyette, Swets, Meijer, Wouters, & Authors, 2010; McKay, Danyko, Neziroglu, & Yaryura-Tobias, 1995) or three (obsessions, compulsions and resistance; Kim, Dysken, Pheley, & Hoover, 1994; Moritz et al., 2002) rather independent factors. However, we need to acknowledge that there are some exceptions, with studies reporting somewhat higher correlations between obsessions and compulsions (e.g., Gönner & Kupfer, in preparation; McKay et al., 1995). Interestingly, internal consistency was not different across groups, showing that this index is not very sensitive to more complex cases of deceit. Overall, the level of internal consistency was in line with prior studies on the OCI-R (Cronbach's alpha = .80–.87; Abramowitz & Deacon, 2006; Foa et al., 2002; Gönner et al., 2007; Huppert et al., 2007; Sica et al., 2009; Sulkowski et al., 2008) and the Y-BOCS self-report scale (Cronbach's alpha = .87 and .78; Gönner & Kupfer, in preparation; Steketee et al., 1996). The correlation between OCI-R and Y-BOCS did not represent a stable discriminator as the simulators were even closer to the values reported by Gönner et al. (2008) than the patient samples. While these three cross-sectional indices for deception ((1) inflated total scores on one scale (OCI-R) but not the other (Y-BOCS), (2) inflated correlation between obsessions and compulsions, and (3) low correlation between OCI-R and Y-BOCS) differentiated samples in the current study, they may not be useful for the identification (and thus exclusion) of individual simulators. However, reporting these indices in future studies may indicate how far the collected data may be trusted.

The results are noteworthy in our view as we chose a rather conservative design by selecting a sample of mental health experts with different levels of expertise on OCD who should be more capable than non-experts of feigning OCD symptoms. In other words, the psychometric properties results of "liars" are hypothesized to be even worse in a naturalistic design and we speculate that "liars" would often either cancel participation or not participate at retest.

To conclude, our study renders it unlikely that online investigations on OCD with good psychometric properties contain many subjects with fictitious disorders. Nonetheless, we cannot rule out that some patients in Internet trials have (additional) diagnoses that would usually lead to exclusion from conventional clinical studies (e.g., severe neurological disorders, certain comorbid axis I or II diagnoses). This objection is graver for basic research studies that usually impose quite rigid selection criteria and enroll "pure" samples. For intervention studies this is less of a problem considering that there is a recent trend in psychopathological research to target "typical" patients (Hollon & Wampold, 2009). Comorbid disorders that would result in exclusion from basic research studies are usually tolerated in intervention trials—non-detection of comorbid disorders in online studies can thus be considered a lesser evil. To guard outcome data against the effects of abuse, we advise researchers to run randomized controlled trials so that participants who try to deceive are most likely equally distributed across conditions. For single armed studies with poor psychometric indices there is a risk that these contain a substantial proportion of subjects who violate inclusion criteria. Whether such subjects are more prone to simulate improvement, deterioration or stable symptomatology is hard to predict, although we speculate that such subjects may rather fake deterioration and are less inclined to take part on multiple occasions.

Our study faces a number of limitations. Firstly, our study did not identify specific subjects who feign symptoms and is also silent on disorders other than OCD. Such studies should be conducted with a large naturalistic online sample and involve specific lie (e.g., items tapping into pseudo-OCD symptoms such as being "obsessed by ghosts") and control scales (e.g., asking the same information twice in disguised form, for example, length of illness (in years) and onset of symptoms (date); age and date of birth; asking about primary symptoms at different stages of the assessment). Secondly, our results should be tested with a non-expert group which may have achieved even worse psychometric properties. Thirdly, future studies should consider other measures in addition to Y-BOCS and OCI-R. Further, we recommend that studies offer non-monetary compensation for participation, for example, a treatment manual, which is not attractive to a non-clinical population.

While we do not deny that results from Internet studies, perhaps more than FTF clinical studies, need multiple replications before solid conclusions can be drawn, we concur with, for example Gosling et al. (2004), that online studies, notwithstanding their alleged weaknesses, are reliable when certain precautions are met. Consideration of psychometric properties may help to assess the quality and thus trustworthiness of an online investigation.

References

Abramowitz, J. S., & Deacon, B. J. (2006). Psychometric properties and construct validity of the Obsessive-Compulsive Inventory-Revised: replication and extension with a clinical sample. *Journal of Anxiety Disorders*, *20*, 1016–1035.

Abramowitz, J. S., Tolin, D., & Diefenbach, G. (2005). Measuring change in OCD: sensitivity of the Obsessive-Compulsive Inventory-Revised. *Journal of Psychopathology and Behavioral Assessment*, *27*, 317–324.

Baer, L., Brown-Beasley, M. W., Sorce, J., & Henriques, A. I. (1993). Computer-assisted telephone administration of a structured interview for obsessive-compulsive disorder. *American Journal of Psychiatry*, *150*, 1737–1738.

Author's personal copy

Besiroglu, L., Cilli, A. S., & Askin, R. (2004). The predictors of health care seeking behavior in obsessive-compulsive disorder. *Comprehensive Psychiatry*, *45*, 99–108.

Boyette, L., Swets, M., Meijer, C., Wouters, L., & Authors, G. R. O. U.P. (2010). Factor structure of the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) in a large sample of patients with schizophrenia or related disorders and comorbid obsessive-compulsive symptoms. *Psychiatry Research*, *186*, 409–413.

Brett, C. M., Johns, L. C., Peters, E. P., & McGuire, P. K. (2009). The role of metacognitive beliefs in determining the impact of anomalous experiences: a comparison of help-seeking and non-help-seeking groups of people experiencing psychotic-like anomalies. *Psychological Medicine*, *39*, 939–950.

Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., et al. (2005). Non-equivalence of online and paper-and-pencil psychological tests: the case of the Prospective Memory Questionnaire. *Behavior Research Methods, Instruments and Computers*, *37*, 148–154.

Coles, M. E., Cook, L. M., & Blake, T. R. (2007). Assessing obsessive compulsive symptoms and cognitions on the internet: evidence for the comparability of paper and Internet administration. *Behaviour Research and Therapy*, *45*, 2232–2240.

Coles, M. E., Johnson, E. M., & Schubert, J. R. (2011). Retrospective reports of the development of obsessive compulsive disorder: extending knowledge of the protracted symptom phase. *Behavioural and Cognitive Psychotherapy*, *39*, 579–589.

Cuijpers, P., Donker, T., Johansson, R., Mohr, D. C., van Straten, A., & Andersson, G. (2011). Self-guided psychological treatment for depressive symptoms: a meta-analysis. *PLoS One*, *6*, e21274.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., et al. (2002). The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychological Assessment*, *14*, 485–496.

Gönner, S., & Kupfer, J. The Yale-Brown Obsessive Compulsive Scale: reliability of the self-report form, in preparation.

Gönner, S., Leonhart, R., & Ecker, W. (2007). Das Zwangsinventar OCI-R-die deutsche Version des Obsessive-Compulsive Inventory-Revised: ein kurzes Selbstbeurteilungsinstrument zur mehrdimensionalen Messung von Zwangssymptomen [The German version of the Obsessive-Compulsive Inventory-revised: a brief self-report measure for the multidimensional assessment of obsessive-compulsive symptoms]. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *57*, 395–404.

Gönner, S., Leonhart, R., & Ecker, W. (2008). The Obsessive-Compulsive Inventory-Revised (OCI-R): validation of the German version in a sample of patients with OCD, anxiety disorders, and depressive disorders. *Journal of Anxiety Disorders*, *22*, 734–749.

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

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analyses of six preconceptions about internet questionnaires. *American Psychologist*, *59*, 95–104.

Hancock, J. T. (2007). Digital deception: why, when and how people lie online. In: A. N. Joinson, K. Y.A. McKenna, T. Postmes, & U. D. Reips (Eds.), *The Oxdford Handbook of Internet Psychology* (pp. 289–301). Oxford: Oxford University Press.

Hollon, S. D., & Wampold, B. E. (2009). Are randomized controlled trials relevant to clinical practice?. *Canadian Journal of Psychiatry*, *54*, 637–643.

Huppert, J. D., Walther, M. R., Hajcak, G., Yadin, E., Foa, E. B., Simpson, H. B., et al. (2007). The OCI-R: validation of the subscales in a clinical sample. *Journal of Anxiety Disorders*, *21*, 394–406.

Kim, S. W., Dysken, M. W., & Kuskowski, M. (1990). The Yale-Brown Obsessive-Compulsive Scale: a reliability and validity study. *Psychiatry Research*, *34*, 99–106.

Kim, S. W., Dysken, M. W., Pheley, A. M., & Hoover, K. M. (1994). The Yale-Brown Obsessive-Compulsive Scale: measures of internal consistency. *Psychiatry Research*, *51*, 203–211.

Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). The treatment gap in mental health care. *Bulletin of the World Health Organisation*, *82*, 858–866.

Marques, L., LeBlanc, N. J., Weingarden, H. M., Timpano, K. R., Jenike, M., & Wilhelm, S. (2010). Barriers to treatment and service utilization in an internet sample of individuals with obsessive-compulsive symptoms. *Depression & Anxiety*, *27*, 470–475.

Mataix-Cols, D., & Marks, I. M. (2006). Self-help with minimal therapist contact for obsessive-compulsive disorder: a review. *European Psychiatry*, *21*, 75–80.

McKay, D., Danyko, S., Neziroglu, F., & Yaryura-Tobias, J. A. (1995). Factor structure of the Yale-Brown Obsessive-Compulsive Scale: a two dimensional measure. *Behaviour Research & Therapy*, *33*, 865–869.

Moritz, S., & Hauschildt, M. (2011). *Erfolgreich gegen Zwangsstörungen: Metakognitives Training-Denkfallen erkennen und entschärfen [Successful against OCD. Metacognitive training-detecting and defusing cognitive traps]*. Heidelberg: Springer.

Moritz, S., Jelinek, L., Hauschildt, M., & Naber, D. (2010). How to treat the untreated: effectiveness of a self-help metacognitive training program (myMCT) for obsessive-compulsive disorder. *Dialogues in Clinical Neurosciences*, *12*, 209–220 available online at: www.uke.de/mymct.

Moritz, S., Meier, B., Kloss, M., Jacobsen, D., Wein, C., Fricke, S., et al. (2002). Dimensional structure of the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS). *Psychiatry Research*, *109*, 193–199.

Moritz, S., Timpano, K.R., Wittekind, C.E., Knaevelsrud, C. Harnessing the web: Internet and self-help therapy for people with obsessive-compulsive disorder and post-traumatic stress disorder. In E. Storch & D. McKay (Eds.), *Handbook of treating variants and complications in anxiety disorders*. Heidelberg: Springer, in press.

Moritz, S., Wittekind, C. E., Hauschildt, M., & Timpano, K. R. (2011). Do it yourself? Self-help and online therapy for people with obsessive-compulsive disorder. *Current Opinion in Psychiatry*, *24*, 541–548.

Schaible, R., Armbrust, M., & Nutzinger, D. O. (2001). Yale-Brown Obsessive Compulsive Scale: sind Selbst-und Fremdrating äquivalent? [Yale-Brown Obsessive Compulsive Scale: are self-rating and interview equivalent measures?]. *Verhaltenstherapie*, *11*, 298–303.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The MINI International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview. *Journal of Clinical Psychiatry*, *59*(Suppl. 20), 22–33.

Sica, C., Ghisi, M., Altoè, G., Chiri, L. R., Franceschini, S., Coradeschi, D., et al. (2009). The Italian version of the Obsessive Compulsive Inventory: its psychometric properties on community and clinical samples. *Journal of Anxiety Disorders*, *23*, 204–211.

Steketee, G., Frost, R., & Bogart, K. (1996). The Yale-Brown Obsessive Compulsive Scale: interview versus self-report. *Behaviour Research and Therapy*, *34*, 675–684.

Sulkowski, M. L., Storch, E. A., Geffken, G. R., Ricketts, E., Murphy, T. K., & Goodman, W. K. (2008). Concurrent validity of the Yale-Brown Obsessive-Compulsive Scale-Symptom Checklist. *Journal of Clinical Psychology*, *64*, 1338–1351.

Whitehead, L. C. (2007). Methodological and ethical issues in Internet-mediated research in the field of health: an integrated review of the literature. *Social Science and Medicine*, *65*, 782–791.

Woody, S. R., Steketee, G., & Chambless, D. L. (1995). Reliability and validity of the Yale-Brown Obsessive-Compulsive Scale. *Behaviour Research and Therapy*, *33*, 597–605.

Wootton, B. M., Titov, N., Dear, B. F., Spence, J., & Kemp, A. (2011). The acceptability of Internet-based treatment and characteristics of an adult sample with obsessive compulsive disorder: an Internet survey. *Plos One*, *6*, e20548.