International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

The Role of Clinical Phenomena Evaluation in Mechanical M Turk Studies

M NARSIMHA 1, POLAMURI RAMA MOHAN REDDY 2,
ASSOCIATE PROFESSOR 2, ASSISTANT PROFESSOR 1,
Mail ID:narsimha329@gmail.com, Mail ID:polamurirammohanreddy@gmail.com
Dept.: Mechanical
Pallavi Engineering College,
Kuntloor(V),Hayathnagar(M),Hyderabad,R.R.Dist.-501505.

Abstract

The Murk platform on Amazon.com enables rapid, low-cost, and demographically representative data collecting. Multiple articles have praised Murk for its ability to collect high-quality data from an epidemiological sample that is more typical of the U.S. population than conventional in-person convenience samples (e.g., undergraduate subject pools). Because of this advantage, as well as the simplicity and cheap cost of data collection, the number of research use MTurk to probe phenomena in a variety of psychological subfields has increased dramatically in recent years. In recent years, several studies have looked at how MTurk samples compare to the general population. However, there is still a major knowledge gap because of the lack of information about the variability of clinical symptoms among Murk participants. This research argues that identifying clinical phenomena in MTurk samples is crucial, and it backs up these claims with data from a large-scale empirical study of MTurk participants (N = 1,098). Compared to typical non-clinical samples, MTurk users strongly endorse clinical symptoms. This difference was particularly pronounced in regards to the endorsement of depressive and social anxiety symptoms, which were at levels similar to those of those with clinically confirmed mood and anxiety disorders. All of the participants' physiological anxiety, hoarding, and eating pathological symptoms were below clinical levels. The prevalence estimates for 12 months were 3–19 times higher among those who met the verified clinical cut-offs. Researchers should be wary of referring to the MTurk sample as typical of the community at large, it is suggested, since Murk participants vary from the broader population in important ways.

Keywords:

Psychopathology, Signs, Exposure, and Incidence on Amazon's Mechanical Turk (MTurk)

introduction

There has been a recent movement in several subfields of psychology toward the use of non-laboratory research techniques to augment studies undertaken in laboratories, and one such way is the use of Mechanical Turk for the assessment of clinical phenomena (e.g., Reis & Gosling, 2010). Accordingly, Amazon. Om's Mechanical Turk (Mturk) website offers a platform through which registered people from throughout the world, referred to as workers, may conduct surveys and/or automated tasks for a modest money reward. Quick, simple, and cheap access to a large and varied sample of persons are just a few of the methodological features that make Mturk so appealing to the research community (for reviews, see Burmaster, Kwang, & Gosling, 2011, and Polacca, Chandler, &Ipe rotis, 2010). Consequently, a vast number of studies in the fields of social psychology, evolutionary psychology, cognitive psychology, emotion research, and clinical psychology have been undertaken utilizing Mturk (Belinsky, Huber, & Lenz, 2012; Mason & Suri, 2012). The demographics of Murk samples have been the subject of several papers (e.g., Behrend, Share, Meade, & Wiebe, 2011; Goodman, Crider, & Cheema, 2013; Polacca et al., 2010), but little is known about individual variations in clinical symptoms among Mturk participants.

This paper aims to shed light on the question of how similar MTurk workers and the general US population really are by discussing the significance of assessing clinical phenomena in MTurk samples and analysing the similarities and differences between the two groups on a variety of clinical characteristics. A first look towards answering the mystery of "who are MTurk workers?" research has shown that Murk samples are more representative of the broader population than either undergraduate or other Internet samples. One study indicated that Mturk employees were older and more varied in terms of ethnicity than undergraduate participants. The MTurk workforce may also be more varied in terms of race, ethnicity, and socioeconomic status than other representative samples of the Internet population (Casper, Bickel, & Hackett, 2013; Gosling, Vizier, Srivastava, & John, 2004). However, there are still some noticeable discrepancies when comparing them to community samples.

One research indicated that Mturk participants were representative of the overall population in terms of gender, education, and age (Goodman et al., 2013), while another reported that their sample of Mturk participants was more feminine and somewhat younger than the general population (Polacca et al., 2010). Even though they make less money, Mturk employees may have more education than the typical American (Ipe rotis, 2010; Polacca et al., 2010; Shapiro, Chandler, & Mueller, 2013). Based on the findings of this study, it seems that Mturk employees vary significantly from community-based participants, despite the fact that Mturk gives researchers access to a sample that is more

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

representative of the general population than other commonly utilized subject pools.

Method

Participants

The data was compiled through a multi-pronged study that included five individual MTurk surveys. Workers from Mturk with a 90% positive rating or above who live in the United States were considered for inclusion. Employees who took part in more than one survey had their data combined, and duplicate entries were eliminated so that only their most recently completed survey answers would be included in the analysis. People who finished the survey in less than 60% of the estimated time to finish it (n = 230) were not included in the analysis, since this is consistent with past Mturk research (e.g., Behrend et al., 2011). There was no statistically significant difference in gender, race, or ethnicity between the groups included and excluded from the analysis (all PS>.10). However, those who were left out were significantly younger than those who were included (M = 27.18, SD = 9.96 years vs. M = 31.16, SD = 13.06), t (1326) = 4.36, p .001. There was a total of 1,098 Mturk employees in the sample (n = 204 for Survey 1, 399 for Survey 2, 333 for Survey 3, 74 for Survey 4, and 88 for Survey 5). Due to the incomplete nature of most surveys, the total number of answers varies across the measurements (see Tables 1 and 3). Over half (51.5%) of the participants were female; the average age was 31.16; the majority (79.0%) were White/Caucasian; 7.9% were Black/African American; 7.3% were Asian; 1.5% were multi-racial; 4.3% classified as "other;" 7.7% were Hispanic/Latino.

Procedure

The [Blinded for Review] Institutional Review Board has seen and approved all research protocols. Before giving their permission, participants read a short explanation of the research. They then responded to a survey that included some of the following indicators. Analyses comparing the present sample to clinical and non-clinical samples employed previously published means, standard deviations, and/or clinical threshold scores for each of these parameters (described in detail below). It took respondents anything from 15 minutes to an hour to finish each survey. Participants were paid between $4 and $10 per hour, depending on the survey, which is on line with several Mturk studies (e.g., Belinsky et al, 2012; Ipe rotis, 2010) but more than the median hourly salary for Mturk jobs that was previously published (Horton & Chilton, 2010).

Indicators of Symptoms

Stress, Anxiety, and Depression Rating Scales (DASS-21; Henry & Crawford, 2005).

The DASS-21's Depression and Anxiety measures were used in this analysis. On a scale from 0 (did not apply to me at all) to 3 (very lot), participants indicated how often they had experienced depressive or physiological anxiety symptoms in the previous week (applied to me very much). Data from a non-clinical sample found that the mean scores on the Depression and Anxiety Subscales were 3.87 (SD = 3.98) and 3.18 (SD = 3.38) (Osman et al., 2012). The average depression and anxiety subscale scores of those who visited an outpatient mental health clinic were 10.65 (SD = 9.30) and 10.90 (SD = 8.12), respectively (Brown, Charita, Kurtotic, & Barlow, 1997). Lovibond and Lovibond (1995) provide the following criteria for the intensity of depressive symptoms, despite the lack of clinical cut-offs for these subscales: scores of 0-9 (normal), 10-13 (mild), 14-20 (moderate), 21-27 (severe), and 28+ (extreme) (extremely severe). Also, the suggested range for anxiety symptom intensity is as follows: 0-7 for "normal," 8-9 for "mild," 10-14 for "moderate," 15-19 for "severe," and 20+ for "intense" (extremely severe).

Scale for Measuring Obsessive Compulsive Disorder (DOCS; Abramowitz et al., 2010).

Measurement of obsessive-compulsive symptoms is provided by the DOCS, a 20-item scale. The items are given a score between 0 and 4, with the anchors altering according on the kind of object being scored. Previous studies have demonstrated that a clinical cut-off of 21 provides the optimal mix between sensitivity and specificity in properly diagnosing OCD, with mean DOCS scores of 11.93 (SD = 9.87) for students and 30.06 (SD = 14.49) for patients with OCD (Abramowitz et al., 2010).

Evaluation of Overeating and Related Disorders (EDI; Garner, Olmstead, &Policy, 1983).

Symptoms of eating disorders may be measured using the Eating Disorders Inventory (EDI). In this analysis, two measures were employed: the "Drive for Thinness" and the "Bulimia" scales. Each participant rates each item on a scale from 1

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

(never) to 6 (very often) across these subscales (always). Following this, ratings between 1 and 3 are recoded as 0, while ratings between 4 and 6 are recoded as 1 to 3. In the past, researchers have shown that female college students had a mean Drive for Thinness score of 5.0 (SE =.22), whereas female anorexic patients have a score of 15.4 (SE =.50). (Garner et al., 1983). Previous research indicated that female college students had mean bulimia scores of 2.0 (SE =.14), whereas female patients with the bulimia subtype of anorexia had mean scores of 10.8 (SE =.69). As far as we are aware, no clinical cut-offs exist for these two subscales.

A Self-Rating Scale for Hoarding Disorder (HRS-SR; Tolan, Frost, &Steele, 2010).

Participants score their hoarding symptoms on a scale from 0 (none) to 8 (very severe) on the Hoarding Rating Scale-Short Version (HRS SR) (extreme). Participants without a psychiatric history scored a mean of 3.34 on the HRS interview (SD = 4.97), whereas individuals with hoarding scored a mean of 6.33.

standard deviation = 5.67; Tolan et al., 2010 mean = 24.22 Clinically, Tolan and co-workers found that a cut-off score of 14 distinguished OCD patients from hoarders the most effectively.

A Measure of Anxiety Regarding Social Interaction (SIAS; Mattock& Clarke, 1998).

The Social Anxiety and Stress Scale (SIAS) is a Likert-scale self-report assessment of social anxiety symptoms (extremely characteristic or true of me). Patients with social phobia had a mean SIAS score of 34.60 (SD = 16.40) compared to a previous study's unselected sample mean SIAS score of 18.80 (SD = 11.80) (Mattock& Clarke, 1998). Based on the research of Peters (2000), the optimal cut-off score for this instrument is 36. This number strikes a good compromise between sensitivity and specificity.

Anxiety Sensitivity Index-3 for Measuring Cognitive Vulnerability (ASI-3; Taylor et al., 2007).

Anxiety-related symptom dread is a well-established risk factor for anxiety disorders, and the ASI-3, an 18-item assessment, assesses this fear (Schmidt, Smolinsky, &Manner, 2006). Everything is ranked on a scale from 0 (very little) to 4 (extremely much) (very much). Previous studies have indicated that individuals with panic disorder had an average ASI-3 score of 32.69 (SD = 15.21), whereas unselected college students score 13.83 (SD = 10.79). (Wheaton, Deacon, McGrath, Berman, & Abramowitz, 2012).Stress Reaction Questionnaire (DTS; Simons &Gather, 2005). An individual's distress tolerance, as measured by the DTS (Simons &Gather, 2005), is a vulnerability factor associated with an increased risk of a wide range of mental health issues (Simons &Gather, 2005). (For a review, see Lepro, Smolinsky, & Bernstein, 2010). The participants are given a set of statements and asked to evaluate how much they agree or disagree with each one using a 5-point scale (1 = strongly disagree, 5 = completely agree) (strongly disagree). An increased capacity for tolerance is indicated by a high mean score.

distress. DTS mean scores have been reported in the literature to be 3.43 (SD =.83) in non-selected undergraduates and 2.87 (SD =.89) in individuals with primary anxiety disorders (Mitchell, Riccardo, Keough, Timpano, & Schmidt, 2013).

Measuring Your Tolerance for Uncertainty (IUS; Carleton, Sharpe, & Asmundson, 2007).

The IUS is a 12-item scale that measures the propensity to have a negative emotional and behavioural response to ambiguous situations (Buhr & Dugas, 2002). Numerous types of anxiety have been linked to an intolerance of uncertainty (for a review: Lepro et al., 2010). Items are rated by respondents between 1 (totally not me) and 5 (very much like me) (entirely characteristic of me). Patients with generalized anxiety disorder had a mean IUS score of 40.38 (SD = 11.26; Carleton et al., 2007), whereas those who did not have an anxiety disorder had a score of 29.53 (SD = 10.96) among Internet-based participants.

Measure of Introspective Reflection (RRS; Treynor, Gonzalez, & Nolen-Hoeksema, 2003).

The RRS is a test for rumination, a maladaptive coping mechanism that entails dwelling on one's unpleasant emotions and the ways in which one's life has changed because of them (Nolen-Hoeksema, Wesco, & Lyubomirsky, 2008). In addition to its links to depressive thoughts and behaviours, ruminating has also been linked to binge eating and anxiety (Nolen Hoeksema et al., 2008). Using a scale from 1 (nearly never) to 4 (very often), participants evaluate how often they

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

tend to focus on specific elements of dysphoric mood while completing the RRS (almost always). Previous research found that whereas individuals with no history of Axis I disorders had a mean RRS score of 29.90 (SD = 7.66), those experiencing a current major depressive episode had a mean RRS score of 59.90 (SD = 14.13; Zetsche, Davanzo, &Doorman, 2012).

Table 1 presents descriptive data for the clinical symptom assessments in our Mturk sample, indicating a high prevalence. The EDI-Bulimia subscale had a positive skew and a leptokurtic distribution, while all other measures showed skewness and kurtosis values in the excellent (1) to acceptable (2) range. Internal consistency was high (0.80 or above) across the board. This study compared the present sample's mean and standard deviation to those of previously published non-clinical and clinical samples to evaluate the existence of clinically relevant symptoms. Table 2 shows the effect sizes and confidence ranges for these contrasts. Those involved generally had clinically substantial levels of both social anxiety and sadness. The effect size and confidence intervals were considerable when comparing the sample mean SIAS score to a previously acquired non-clinical mean, but minor when comparing the sample mean to a previously obtained clinical mean. The trend was also seen when comparing the mean DASS-Depression score in the present group to non-clinical and clinical averages obtained in the past. Participants' mean DASS-Anxiety scores showed a moderate-to-large effect size when compared to pre-study non-clinical means. However, a somewhat significant impact was also seen when this mean was compared to a prior clinical mean. The physiological anxiety symptoms reported by Mturk participants seem to have been below the threshold for clinical diagnosis. Subclinical levels of eating disorders were also reported by Murk participants.

subscale of the Eating Disorder Inventory (EDI) called "Drive for Thinness," and hoarding symptoms on the Hoarding Rating Scale for DSM-IV. Both the mean DOCS and EDI-Bulimia values for the sample were outside of clinical territory. When comparing our Mturk sample's symptom ratings to both historical non-clinical means and clinical means, we found very tiny to small impacts with confidence intervals that included zero and very large significant effects, respectively. The proportion of those whose symptoms were at or above the clinical criteria for each metric was then analysed. Figure 1 displays these percentages, rounded to the closest whole number, with the 12-month prevalence rates that have previously been reported for each disease. The proportion of individuals with clinical levels of social anxiety is seven times higher than the expected 12-month prevalence rate, as seen in the figure. Furthermore, the projected 12-month prevalence rates for mood disorders (which include unipolar and bipolar disorders) and panic disorder were almost 3x and 9x, respectively, the proportion of individuals with at least moderate levels of sadness and anxiety. As a conclusion, the percentage of participants reporting clinical levels of obsessive-compulsive symptoms was 19 times the estimated 12-month prevalence rate for OCD, and the percentage of individuals reporting clinical levels of hoarding symptoms was almost 7 times the 12-month prevalence rate for hoarding disorder.

Signs of Cognitive Weakness Exist

Table 3 presents descriptive data for cognitive vulnerability assessments. All four measurements have respectable skewness and kurtosis values (between -1 and +1). Furthermore, there was a high degree of consistency across the four different assessments (.84 or above) among researchers who used them. Participants.

showed increased susceptibility to cognitive breakdown in all areas. The sample means for all the parameters were above the non-clinical means that had been established before. The sample means for the IUS and RRS were more closely aligned with the established clinical means. Percentages ranging from 21.7% to 37.4% were found to be beyond the clinical mean (or outside the DTS's clinical minimum, respectively).

Correlations by Means of Chronological Age, Sexual Orientation, and Race/Ethnicity

The prevalence of several clinical symptoms was lower among older individuals, while the effect sizes were minor. Age was shown to be inversely related to scores on the DASS for both depression and anxiety (r = -.16, p =.001), the DOCS (r = -.12, p =.001), and the SIAS (r = -.17, p =.001). Similar to the negative relationships between age and ASI-3 (r = -.16, p =.001) and RRS (r = -.16, p =.003) scores, there was a positive connection between age with DTS scores (r =.12, p =.03), suggesting that older individuals endorsed less cognitive vulnerability. Means of symptoms and cognitive vulnerability assessments by gender are shown in Table 4. Women were found to have a higher prevalence of reporting symptoms of eating disorders and social anxiety than males. On the Anxiety Severity Index (ASI-3), women reported

higher scores than males, suggesting that they, on average, advocate more dread of anxiety symptoms. Fewer noteworthy results were found in analyses that focused on racial and ethnic groups. OCD symptoms were rated as less severe among whites (M = 12.92, SD = 10.81) than those of people of color (M = 15.25, SD = 11.25), t (791) = 2.43, p =.02, d =.21. Furthermore, Hispanic individuals reported higher ASI-3 scores (M = 28.76, SD = 18.61) than non-Hispanic participants (M = 20.71, SD = 14.83), t (285) = 2.35, p =.02, d =.48.

In addition, no statistically significant variations were discovered in any of the symptom or cognitive vulnerability assessments.

conclusion

Psychological symptoms and cognitive vulnerability variables were among the clinical features of a large Mturk population that were investigated in this research. Participants reported high levels of social anxiety symptoms, with over half receiving a SIAS score above the suggested clinical threshold, which is in line with earlier research (Shapiro et al., 2013). Although the sample mean on the DASS Depression subscale was not substantially different from a previously obtained clinical mean, 32% of the sample indicated at least moderate degrees of depression, which contradicts prior research (Brown et al., 1997). In terms of physiological anxiety, the desire to be skinny, and hoarding symptoms, participants reported subclinical levels, as shown by scores that were greater than previously reported non-clinical means but lower than previously reported clinical means. In contrast to typical population or epidemiological samples, Mturk employees are more likely to exhibit a number of psychological disorders, including social anxiety and sadness. No more Mturk employees reported having symptoms of bulimia or OCD than would be expected based on data from the general population. Even so, it's worth noting that 19% of the sample reached a DOCS score over the clinical limit; this is a far higher proportion than is normally found in epidemiological studies (Kessler et al., 2005). Finally, as predicted, individuals reported higher levels of cognitive vulnerability than the overall population.

References

[1] Abramowitz, J. S., Deacon, B. J., Olatunji, B. O., Wheaton, M. G., Berman, N. C., Losar do, D., Timpano, K. R., et al. (2010). Assessment of obsessive-compulsive symptom dimensions: development and evaluation of the Dimensional Obsessive-Compulsive Scale. Psychological Assessment, 22, 180-198. doi:10.1037/a0018260

[2] Behrend, T. S., Share, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. Behavioural Research Methods, 43, 800-813. Doi: 10.3758/s13428-011-0081-0

[3]Belinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labour markets for experimental research: Amazon. Om's Mechanical Turk. Political Analysis, 20, 351-368. Doi: 10.1093/Pan/Mpr057

[4] resale, J., Kendler, K. S., Us, M., Gaviola-Aguilar, S., & Kessler, R. C. (2005). Lifetime risk and persistence of psychiatric disorders across ethnic groups in the United States. Psychological Medicine, 35, 317-327. Doi: 10.1017/S0033291704003514

[5] Brown, T. A., Charita, B. F., Kurtotic, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples.

[6] Behaviour Research and Therapy, 35, 79-89. Doi: 10.1016/S0005-7967(96)00068-X Buhr, K., & Dugas, M. J. (2006). Investigating the construct validity of intolerance of uncertainty and its unique relationship with worry. Journal of Anxiety Disorders, 20, 222-236. Doi: 10.1016/j.janxdis.2004.12.004.

[7] Burmaster, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? Perspectives of Psychological Science, 6, 3-5, Doi: 10.1177/1745691610393980

[8] Carleton, R. N., Sharpe, D., & Asmundson, G. J. G. (2007). Anxiety sensitivity and intolerance of uncertainty: requisites of the fundamental fears? Behaviour Research and Therapy, 45, 2307-2316. doi: 10.1016/j.brat.2007.04.006

[9] Casper, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's Murk, social media, and face-to-face behavioural testing. Computers in Human Behaviour, 29, 2156-2160. Doi: 10.1016/j.chb.2013.05.009

[10] Garner, D. M., Olmstead, M. P., &Policy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia-nervosa and bulimia. International Journal of Eating Disorders, 2, 15-34.

[11] Goodman, J. K., Crider, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. Journal of Behavioural Decision Making, 26, 213-224. Doi: 10.1002/bdm.1753

[12] Gosling, S. D., Vizier, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. American Psychologist, 59, 93-104. Doi: 10.1037/0003-066X.59.2.93

[13] Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. British Journal of Clinical Psychology, 44, 227-239. doi:10.1348/014466505X29657

[14] Horton, J. J., & Chilton, L. B. (2010). The labour economics of paid crowdsourcing. Paper presented at the Proceedings of the 11th ACM conference on electronic commerce.

[15] Ipe rotis, P. G. (2010). Analysing the amazon mechanical Turk marketplace. XRDS: Crossroads, The ACM Magazine for Students, 17, 16-21. Doi: 10.1145/1869086.1869094

[16] Doorman, J., Livens, S. M., &Gottlieb, I. H. (2011). Sticky thoughts: Depression and rumination are associated with difficulties manipulating emotional material in working memory. Psychological Science, 22, 979-983. Doi: 10.1177/0956797611415539

[17] Jorum, A. F. (2000). Does old age reduce the risk of anxiety and depression? A review of epidemiological studies across the adult life span. Psychological Medicine, 30, 11-22. Doi: 10.1017/S0033291799001452

[18] Kessler, R. C., Chiu, W. T., Delmer, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry, 62, 617-627. Doi: 10.1001/archpsyc.62.6.617