HHS Public Access

Author manuscript

J Trauma Stress. Author manuscript; available in PMC 2019 August 01.

Published in final edited form as:

J Trauma Stress. 2018 August ; 31(4): 549–557. doi:10.1002/jts.22303.

Crowdsourcing Trauma: Psychopathology in a Trauma-Exposed Sample Recruited via Mechanical Turk

Katherine van Stolk-Cooke[1], **Andrew Brown**[1], **Anne Maheux**[1], **Justin Parent**[1], **Rex Forehand**[1], and **Matthew Price**[1]

[1]Department of Psychology, University of Vermont, Burlington, Vermont, USA

Abstract

Although crowdsourcing websites like Amazon's Mechanical Turk (MTurk) allow researchers to conduct research efficiently, it is unclear if MTurk and traditionally recruited samples are comparable when assessing the sequela of traumatic events. We compared the responses to validated self-report measures of posttraumatic stress disorder (PTSD) and related constructs that were given by 822 participants recruited via MTurk and had experienced a *DSM-5* Criterion A traumatic event to responses obtained in recent samples of participants recruited via traditional methods. Results suggested that the rate of PTSD in the present sample (19.8%) was statistically higher than that found in a recent systematic review of studies that used only traditional recruitment methods. The severity of PTSD reported in the MTurk sample was significantly greater than that reported in a college sample, $d = 0.24$, and significantly less than that reported in a veteran sample, $d = 0.90$. The factor structure of PTSD found in the MTurk sample was consistent with prevailing models of PTSD. Findings indicate that crowdsourcing may improve access to this hard-to-reach population.

Exposure to potentially traumatic events (PTEs) increases risk for psychiatric disorders, including posttraumatic stress disorder (PTSD). Approximately 6.8% of adults in the United States experience PTSD (National Institutes of Mental Health, 2016). Posttraumatic stress disorder is associated with significant impairment, with disability persisting even when symptoms remit (Bryant et al., 2016). Given the public health burden imposed by PTEs, continued research on PTSD and other trauma-related conditions is needed; however, PTSD includes symptoms such as avoidance of reminders of the PTE and social withdrawal (National Institutes of Mental Health, 2016). These symptoms are barriers to recruitment for research on PTE-related psychopathology, and strategies to overcome these barriers are needed.

Individuals with PTSD may be reluctant to attend in-person research studies that require disclosure of trauma-related information. Relatedly, stigma impedes disclosure of information about traumatic experiences or symptoms (Miller, Canales, Amacker, Backstrom, & Gidycz, 2011). As such, traditional recruitment methods may miss individuals who are unwilling to discuss trauma symptoms in person—a potentially significant portion

Correspondence concerning this article should be sent to Katherine van Stolk-Cooke, University of Vermont, Department of Psychology, John Dewey Hall, 2 Colchester Avenue, Room 230, Burlington, VT 05405. kvanstol@uvm.edu.

of PTE-exposed research candidates. What's more, traditional media recruitment strategies, such as television, radio, and newspaper advertisements, are often expensive and inconsistent (Garrett et al., 2000). Researchers may instead recruit from clinical services, such as emergency departments or specialty clinics that serve this population (Malcoun et al., 2010; Price et al., 2014). People recruited via these avenues often have competing priorities (e.g., pressing clinical or functional needs) that prevent research participation (Price et al., 2017; Rothbaum et al., 2012). Additionally, recruitment strategies for in-person research are restricted by the geographical location in which the research is being conducted, which limits demographic diversity. The academic institution where the present study was conducted is located in a sparsely populated region of the United States that has limited racial/ethnic and socioeconomic diversity. The barriers of these methods have historically limited the scope of research on traumatic stress.

The Internet provides an alternative for recruiting samples for psychological research (Batterham, 2014; Ramo & Prochaska, 2012) that addresses many of the aforementioned barriers (i.e., stigma, burden, geographical limitations). Internet access among adults in the United States increased from 52% in 2000 to 88% in 2016, and 77% of adults in the United States presently own a mobile device with Internet capabilities (Pew Research Center, 2017a, 2017b). These rates of Internet usage are seen across socioeconomic groups, making it a promising method for connecting with underserved groups that are at higher risk for exposure to PTEs (Roberts, Gilman, Breslau, Breslau, & Koenen, 2011).

Research spanning the past several decades has suggested that digital technologies may offer superior methods for collecting data on sensitive or stigmatized information relative to in-person interviews (Kang & Gratch, 2010; Skinner & Allen, 1983; Turner et al., 1998). Jones and colleagues (2014) found that the internet was a preferred medium for information-seeking on health topics due to time efficiency and ease of use. These findings suggest that conducting mental health research on trauma via digital technologies could eliminate the interpersonal barriers associated with discussing a PTE or its sequelae.

Over the last decade, social scientists have begun to use Internet-based crowdsourcing platforms, such as Amazon's Mechanical Turk (MTurk), for experimental and survey-based research on human participants (Huff & Tingley, 2015). MTurk is a platform on which individuals called "requesters" publish jobs, referred to as Human Intelligence Tasks (HITs), and pay individuals called "workers" to complete them. Crowdsourcing is a low-cost, time-efficient, and user-friendly means of accessing a large number of research participants; this has fueled its use for recruitment. The number of social science publications using MTurk data grew from fewer than 50 in 2011 to nearly 550 in 2015 (Chandler & Shapiro, 2016). This influx in MTurk studies has led to questions about the representativeness of MTurk versus in-person samples. Past research has found that MTurk samples are more demographically representative, albeit younger, than convenience samples recruited in person (Berinsky, Huber, & Lenz, 2012; Huff & Tingley, 2015).

The question of representativeness is a key barrier to the adoption of MTurk for mental health related-research. Although there has been an increase in the use of the Internet for research in other social sciences, MTurk's use in mental health research has lagged. Between

2011 and 2015, the number of publications in clinical psychology and psychiatry that used MTurk rose from zero to 50 (Chandler & Shapiro, 2016). Few studies have examined the representativeness of mental health data collected via this method. In the first study to do so, Shapiro, Chandler, and Mueller (2013) concluded that MTurk is a feasible means to evaluate samples with clinically relevant psychopathology and found that prevalence rates of clinically relevant depression symptoms, drug abuse, and PTE exposure were consistent with rates found in the general population. Further, the quality of the authors' MTurk data was high (i.e., high internal validity on psychodiagnositc assessments, comparable associations between psychopathology and demographics to those of in-person samples). This suggests that MTurk respondents may have rates of psychopathology similar to those found in the general population. A thorough examination of the prevalence of trauma exposure and the phenomenology of its associated psychological disorders in an MTurk sample is called for.

There are several metrics by which to evaluate the quality of data collected from participants recruited via novel methods (Chandler, Mueller, & Paolacci, 2014; Chandler & Shapiro, 2016; Paolacci, Chandler, & Ipeirotis, 2010). First, the prevalence of the constructs of interest should be comparable. National estimates of PTSD in individuals who have experienced a PTE range from 6.2–22.9% (Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995; Kilpatrick et al., 2013). Second, the factor structure of the constructs of interest should be the same, as this would suggest that symptom assessments obtained from a trauma-exposed MTurk sample are similar to those obtained via other methods. Factor analyses have supported the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; American Psychiatric Association [APA], 2013) organization of PTSD symptoms into four clusters: intrusions, avoidance, negative alterations in cognitions and mood (NACM), and alterations in arousal and reactivity (AAR; King, Leskin, King, & Weathers, 1998). More recently, a seven-factor hybrid model that includes clusters for intrusions, avoidance, negative affect, anhedonia, externalizing, anxious arousal, and dysphoric arousal was proposed. Finally, previously found associations between the construct of interest and demographic as well as other factors (e.g., related posttrauma psychopathology) should be observed. In prior work, researchers have found high comorbidity between common posttraumatic psychopathologies (i.e., PTSD, depression, and generalized anxiety disorder [GAD]) and that women tend to report more severe posttraumatic psychopathology than men (APA, 2013). Results from several studies have highlighted the comorbidity between these conditions (Price & van Stolk-Cooke, 2015).

In the present study, we examined whether data collected via a large MTurk sample of PTE-exposed individuals was comparable to that obtained through traditional methods. The MTurk sample was considered comparable and representative of the general PTE-exposed sample if (a) the prevalence of likely PTSD and related psychopathologies was similar to those repoted in other studies; (b) the factor structure of PTSD (*DSM-5* and hybrid models) was replicated; (c) expected associations between PTSD scores and demographic factors were found; and (d) expected associations between PTSD and other psychopathology, including depression and GAD, were found, as they so frequently co-occur (Friedman, Resick, Bryant, & Brewin, 2011). Variables relevant to crowdsourcing, (i.e., time taken to complete an assessment, time needed to collect the desired sample size, number of valid

cases) were assessed to determine the number of MTurk participants required to obtain an adequate clinical sample.

Method

Participants and Procedure

Participants were recruited through six MTurk HITs seeking participants to complete questionnaires assessing the impact of stressful events on individuals' lives with the following keywords: *survey, stress, gender, women/men*, and *health*. Participants could access the HIT only if they had a United States–based internet protocol (IP) address and a 65% approval rating from other requesters for prior HITs. Participants were given 1 hr to complete all measures and were compensated $1.25 (USD) for approved work. MTurk provides usage data for HITs (e.g., timestamps for HIT publication and recruitment completion, average HIT completion time). Inclusion criteria for the study included exposure to a Criterion A traumatic event, defined as direct exposure to one of the events on the Life Events Checklist–5 (LEC-5; Weathers, Blake et al., 2013). Individuals who did not endorse a PTE on the LEC-5 were notified of their ineligibility, barred from proceeding or reopening the HIT through IP blocking, and were not compensated. This study was approved by the University of Vermont Institutional Review Board for Human Subjects Research.

Measures

Traumatic event exposure—The Life Events Checklist-5 (LEC-5; Weathers, Blake et al., 2013) is a 17-item self-report measure that assesses exposure to PTEs across the life span. Participants are asked to endorse exposure to 16 known events, with an item included to assess exposure to other extraordinary stressful events. We administered the extended version of the LEC, in which participants were asked to describe the worst event that had happened to them and provide additional information about the circumstances under which it occurred.

PTSD symptoms—The PTSD Checklist-5 (PCL-5; Weathers, Litz et al., 2013) is a 20-item self-report measure that assesses *DSM-5* PTSD symptoms the participant has experienced in the last month. Items assess symptoms across four symptom clusters of PTSD, including intrusions, dysphoria, NACM, and AAR, on a Likert-type scale of 0 to 4. Total scores range from 0–80, with a score of 33 or greater indicating likely PTSD. The PCL-5 was anchored to the most relevant trauma on the LEC. Internal consistency was Cronbach's $\alpha = .95$.

Depression—The Patient Health Questionnaire-8 (PHQ-8; Kroenke, Spitzer, & Williams, 2001) is an eight-item self-report measure that assesses depression symptoms experienced in the last two weeks. Participants rate symptom frequency on a scale of 0 to 3. Scores range from 0 to 24 with higher scores indicating more severe depression. Consistent with prior work, the present study employed a cutoff score of 10 to differentiate individuals with moderate to severe symptomology from those with little to no symptomology (Kroenke et al., 2001; Kroenke et al., 2009). The PHQ-8 is identical to the PHQ-9 except for the removal

of an item on suicidal ideation. Given the intrinsically anonymous nature of MTurk research, it was deemed inappropriate to collect information on ideation/risk from participants upon which the clinical research team could not follow up to assess for safety. Internal consistency was Cronbach's α = .91.

Anxiety—The Generalized Anxiety Disorder-7 (GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006) is a seven-item self-report measure that assesses GAD symptoms the participant has experienced in the last two weeks. Ratings are made on a Likert scale from 0 to 3, with scores ranging from 0 to 21. Higher scores correspond to more severe symptoms. Validation studies have suggested a score of 10 or higher is indicative of severe GAD symptoms (Löwe et al., 2008; Spitzer et al., 2006). Internal consistency was Cronbach's α = .92.

PTSD prevalence rates—We obtained estimates of PTSD prevalence from two sources. The first was a systematic review of 35 longitudinal studies examining rates of PTSD within 12 months after a PTE (Santiago et al., 2013). In this review, authors selected studies that used validated measures capturing *DSM-5* criteria for PTSD, yielding a mean 12-month prevalence of 17.0%. None of the included studies used online methods for data collection. To compare the prevalence of PTSD in the present study to that found in another online sample, we used data published by Kilpatrick and colleagues (2013). Using *DSM-5* criteria, the 12-month prevalence was 8.3%. Estimates of comorbid conditions that result from trauma exposure were obtained from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC; Pietrzak, Goldstein, Southwick, & Grant, 2011). This is the most recent in-person survey to assess PTSD and its comorbid disorders using a large community-based sample of adults. The prevalence rates of PTSD, depression, and GAD were 6.4%, 17.3%, and 8.3%, respectively.

Data Analysis

We conducted primary analyses in SPSS (Version 24). Examination revealed that all variables met assumptions of normality. A series of comparisons using analyses of variance (ANOVAs) with α = .05 were conducted to determine whether any relevant differences existed on variables of interest by demographic characteristics (e.g., gender). We used chi-square tests to compare the prevalence of likely PTSD, depression, and anxiety in our sample to those found in other studies, and relevant differences in symptom severity were assessed using independent samples *t* tests. Associations between PTSD, depression, and GAD were examined using bivariate correlations.

Confirmatory factor analyses (CFAs) were conducted using MPlus (Version 7.11; Muthén & Muthén, 2012) to assess the fit of the factor structure of PTSD symptoms across the *DSM-5* and hybrid models, respectively. Factor variances were scaled to 1 and residual covariances were set to zero. Models were estimated using a polychoric covariance matrix, robust weighted least squares estimation with a mean-and-variance adjusted chi-square (WLSMV) and probit regression coefficients (Flora & Curran, 2004). Model fit was evaluated per Hu and Bentler's (1999) recommendations. Excellent fit was defined as having a comparative fit index (CFI) and Tucker Lewis Index (TLI) ≥ .95 and a root mean square error of

approximation (RMSEA) value ≤ .06. Adequate fit was defined as having a CFI and TLI ≥ .90 and a RMSEA value ≤ .10. Model comparisons were conducted using the Bayesian information criterion (BIC) and maximum likelihood with a robust estimator (MLR) estimation for categorical variables, as this information is not obtained with WLSMV. A 10-point difference in BIC provides very strong support that the model with the lower BIC value fits best (Kass & Raftery, 1995).

Results

We published six HITs to obtain complete data on 600 men and 600 women. As screening occurs within each HIT, more than 1,200 individuals had initial access. MTurk does not count people who screen out and do not complete the survey towards the 1,200 total. As a result, an additional 241 individuals accessed the HIT and were included in the overall access number. The average completion time per assessment was 20 min.

A total of 1,441 individuals accessed the HITs. A small number of participants (n = 126; 8.7%) were excluded for accessing the incorrect HIT for their gender. We assessed the validity of responses with branching logic questions and the inclusion of five questions that asked participants to select a specific response (e.g., “For this item, please select ‘Extremely’”). These items were embedded within the measures as an attention assessment. A valid case was identified as a participant who answered the majority (more than three) of these validity questions correctly and spent more than 5 min completing all measures. This approach is consistent with recommendations to improve the validity of responses on MTurk (Chandler et al., 2016). The responses of an additional 250 (17.3%) participants were deemed invalid and were excluded (23 of these exclusions were for validity failure only). Participants who answered more than three validity questions correctly did not differ from the remaining sample in their PTSD, $t(812) = 0.012$, $p = .761$; depression, $t(802) = 0.379$, $p = .576$; or anxiety symptoms, $t(801) = -1.048$, $p = .455$.

Participants provided a description of their most traumatic experience as part of the extended form of the LEC-5. The descriptions of the remaining 1,065 participants (after exclusions) were reviewed by a licensed clinical psychologist, a clinical graduate student, and a trained bachelors-level research assistant to determine if their experiences met Criterion A for the *DSM-5* PTSD diagnosis. Agreement across the raters was 100%. A total of 243 (16.9%) participants reported events that did not satisfy Criterion A (e.g., loss of a loved one in a manner that was not unexpected or violent, stressful events unassociated with a threat of injury or death). A final sample size of 822 was obtained, constituting 57.0% of individuals who attempted the HIT. Participants ranged in age from 18–82 years.

Demographic variables of the sample are presented in Table 1. Consistent with other MTurk studies, the current sample was younger and had a higher level of education than the general United States population (Huff & Tingley, 2015; Paolacci et al., 2010). The sample was more ethnically diverse than prior MTurk samples in that 73.8% identified as White and non-Hispanic (Paolacci et al., 2010; Shapiro et al., 2013). Participants endorsed a wide array of PTEs (Table 2). The most commonly endorsed PTEs were motor vehicle accidents (61.1%), exposure to natural disasters (39.4%), and physical assault without a weapon

(37.8%). Of participants, 19.3% reported sexual assault and 31.3% reported another unwanted sexual experience.

Descriptive information and correlations among the assessed constructs are presented in Table S1. As expected, PTSD and depression scores, $r = .62$, $p < .001$; PTSD and GAD scores, $r = .62$, $p < .001$; and depression and GAD scores were correlated, $r = .84$, $p < .001$. There was considerable variability in the symptom severity reported by participants for PTSD, depression, and anxiety as well as significant correlations among all three disorders (see Table S1). The prevalence of depression, $\chi^2(1, N = 743) = 83.06$, $p < .001$, and GAD, $\chi^2(1, N = 744) = 204.67$, $p < .001$, were significantly higher than the prevalence rates observed in prior PTE-exposed samples (Pietrzak et al., 2011).

Using established cutoff criteria for each measure, 19.8% of participants reported clinically relevant PTSD symptoms (score greater than 33 on the PCL-5), 27.7% of participants reported moderate to severe depression symptoms (score greater than 10 on the PHQ-8), and 21.5% reported moderate to severe GAD symptoms (score greater than 10 on the GAD-7). Of participants who provided information on symptom severity, and using the aforementioned cutoffs for likely clinical diagnosis, 60.5% of the sample did not meet criteria for any diagnosis, 14.9% had a single likely diagnosis, and 24.6% of the sample had likely comorbid conditions. Of the total sample, 15.8% met criteria for comorbid PTSD and depression, 13.0% met criteria for comorbid PTSD and GAD, and 12.3% met criteria for all three diagnoses.

The prevalence estimate of likely PTSD in the current sample (19.8%) was statistically higher than that which was reported (17.0%) in a recent systematic review by Santiago and colleagues (2013), $\chi^2(1, N = 741) = 4.67$, $p = .031$ (see Table S2). It was also larger than that which was reported for the other online-recruited sample (8.3%; Kilpatrick et al., 2013), $\chi^2(1, N = 741) = 143.57$, $p < .001$ (Table S2). The severity of PTSD reported in the MTurk sample ($M = 19.40$, $SD = 17.73$, 95% CI [18.13, 20.68]) was significantly greater than that which was found in a college sample ($M = 15.42$, $SD = 14.72$, 95% CI [13.72, 17.19]), $t(742) = 6.13$, $p < .001$, and significantly less than that which was found in a veteran sample ($M = 36.97$, $SD = 21.16$, 95% CI [35.04, 38.89]), $t(742) = -26.99$, $p < .001$ (Blevins, Weathers, Davis, Witte, & Domino, 2015; Bovin et al., 2016; see Table S3).

When we compared symptom severity across men and women (see Table S4), we found that women ($M = 20.94$, $SD = 18.89$) reported more severe PTSD symptoms than men ($M = 17.69$, $SD = 16.28$), $F(1, 733) = 6.24$, $p = .013$, both overall and in the symptom clusters of intrusions, $F(1, 733) = 10.92$, $p = .001$, and avoidance, $F(1, 733) = 13.38$, $p < .001$. There were no significant differences between men and women on the PTSD symptom clusters of NACM, $F(1, 733) = 1.21$, $p = .272$, or AAR, $F(1, 733) = 1.61$, $p = .205$. There were no significant differences between women ($M = 6.95$, $SD = 6.18$) and men ($M = 6.18$, $SD = 5.64$) on depression symptoms overall, $F(1, 734) = 3.06$, $p = .081$, or in the affective symptom cluster, $F(1, 723) = 0.73$, $p = .394$. However, there was a significant difference between men and women on somatic depression symptoms, $F(1, 717) = 4.93$, $p = .027$. There were no significant differences between women ($M = 6.18$, $SD = 5.65$) and men ($M = 5.62$, $SD = 5.36$) on GAD symptoms, $F(1, 734) = 1.88$, $p = .171$.

The factor structure of PTSD was evaluated for the *DSM-5* and hybrid models using confirmatory factor analysis (CFA). Factor loadings for each model are presented in Table 3. For the *DSM-5* model, items were loaded on four factors: Intrusions (Items 1–5), Avoidance (Items 6–7), NACM (Items 8–14), and AAR (Items 16–20). For the hybrid model, items were loaded on seven factors: Intrusions (Items 1–5), Avoidance (Items 6–7), Negative Affect (Items 8–11), Anhedonia (Items 12–14), Externalizing (Items 15–16), Anxious Arousal (Items 17–18), and Dysphoric Arousal (Items 19–20). Both the *DSM-5* model, $\chi^2(164, N = 741) = 883.711$, $p < .001$, CFI = .973, TLI = .969, RMSEA = .077; and the hybrid model, $\chi^2(149, N = 741) = 505.383$, $p < .001$, CFI =.987, TLI = .903, RMSEA = .057, fit the data well. A comparison of the Bayesian information criterion (BIC) for each model suggested that the hybrid model best fit the data as indicated the by the lowest overall value (*DSM-5*, BIC = 37822.235; Hybrid, BIC = 37507.279). This suggests a comparable factor structure of PTSD from the current study to that of other studies (Armour et al., 2015; Elhai & Palmieri, 2011; Pietrzak et al., 2015).

We evaluated the factor structure of major depression per the *DSM-5* (Table S5). Items were loaded onto two factors: Affective (Items 1, 2, and 6) and Somatic Symptoms (Items 3, 4, 7, and 8). This model fit the data well, $\chi^2(19, N = 743) = 230.895$, $p < .001$, CFI = .903, TLI = .857, RMSEA = .123, and was comparable to data reported in previously published work (Price & van Stolk-Cooke, 2015).

Discussion

The present study examined the degree to which data on PTE collected via MTurk was comparable to that obtained via traditional methods. Results suggested the prevalence rate of PTSD in the MTurk sample was comparable to those observed in epidemiologic studies of PTE-exposed samples (Santiago et al., 2013) and larger than those obtained through alternative online methods (Kilpatrick et al., 2013). The level of PTSD severity was significantly greater than that which was found in a collegiate sample (Blevins et al., 2015) and less than that of a veteran sample (Bovin et al., 2016). This suggests that MTurk may allow access to individuals with a more severe reaction to PTEs than has been seen in other convenience samples, such as those composed of undergraduates. However, it was expectedly less than that which was found among a population exposed to more severe PTEs (i.e., veterans). Results of a CFA suggest that the factor structures of PTSD according to the *DSM-5* and hybrid models were replicated in the MTurk sample, with both models fitting the data well. The comparable fits of each structural model provide further support as to the utility of MTurk to obtain data that is generalizable to the other PTE-exposed samples. The associations between PTSD symptoms and demographic factors were as expected. We found that PTSD symptom severity was strongly correlated with both depression and GAD, and women reported more severe PTSD symptoms than men. Taken together, these findings suggest that MTurk samples are comparable to those recruited via traditional methods.

The prevalence of PTSD in the MTurk sample was clinically comparable to that which was reported in a systematic review of literature in which traditional recruitment methods had been used, speaking to the promise of crowdsourcing to provide large, reliable samples for research on the effects of trauma exposure with a high degree of efficiency (Santiago et al.,

2013). The obtained responses on measures of PTSD were consistent with previous work in PTE-exposed samples that recruited participants through telephone (Walsh et al., 2012) and in-person interviews (Breslau, Troost, Bohnert, & Luo, 2013). Internal consistency was high for all of the administered measures. That the prevalence rate of PTSD observed in the present study was approximately twice that of the rate observed in Kilpatrick and colleagues' (2013) prevalence report was surprising given the similarities in the online recruitment platforms employed in both cases and the large sample sizes used in each. As such, the discrepancy in PTSD prevalence may be a function of differences in the measures used to assess PTSD or the recruitment methodologies used. Specifically, Kilpatrick and colleagues (2013) stratified their sample according to United States census demographics, which may have affected prevalence estimates.

There are several possible explanations for the elevated estimates of likely depression and GAD. Users of MTurk typically come from more diverse backgrounds with a lower socioeconomic status (SES) than participants in convenience samples, which was confirmed in the current study. As a strong correlate of most psychopathologies, SES may have accounted for the differences in rates of depression and GAD (APA, 2013). In addition, comorbidity between PTSD, depression, and GAD is common, with recent research purporting that symptom overlap may be best understood as an expression of a broader factor of internalizing distress (Kotov et al., 2017). Any examination of a trauma-exposed sample intrinsically selects for individuals who will have experienced distress and are therefore at higher risk for experiencing all three forms of psychopathology. Alternatively, these differences may be attributed to measurement error due to the use of cutoff scores on psychometrically validated self-report instruments as opposed to gold standard interviews.

The MTurk methodology is highly efficient relative to more traditional methods used in survey research. This approach was also cost effective in that 1,441 participants were compensated $1.25 (USD) each, and Amazon charged a service fee of $25 (USD) for a total cost of $1826.25 (USD). Adjusting for the reduced sample size of valid cases, the total cost per participant was approximately $2.22 (USD). Collecting similar data from participants in person or over the telephone would likely require more compensation. Furthermore, the entire study was administered by two individuals. It is unlikely a sample of this size could be acquired at that cost by two research assistants in a reasonable amount of time. The HITs for this study were published in waves over the course of several months, making it unfeasible to compute the exact amount of time taken to complete the study. However, the time taken to republish the survey for each HIT is negligible; thus, it is likely this takes less time than is required to create and administer the individual paper packets that are generally part of more traditional recruitment methods.

Although racial and ethnic diversity in the present sample was limited, the present MTurk sample was more demographically representative than would otherwise have been possible given the geographic location in which this work was conducted. In the present study, 73.8% of participants self-identified as White. United States census data for Vermont, the state in which the research team for the present study resides, has a population of which 95.4% of residents identify as White (Blackwell, Lucas, & Clarke, 2014). Thus, it would be highly unlikely that such a sample would be collected through in-person methods in our geographic

region. Alternative strategies, including telephone-based surveys, would have proven costly, and it is unclear that they would have yielded comparable ethnic proportions. It should be noted, however, that the ratio of African Americans and Latinos in the obtained sample was less than that of current United States census estimates. Given that these groups are disproportionately exposed to PTEs and underrepresented in research (Roberts et al., 2011), scientists should consider the ramifications of relying on crowdsourced data to draw conclusions about trauma-related psychopathology in these individuals. Dedicated work is needed to further determine if MTurk and related approaches are appropriate means of recruiting samples of groups traditionally underrepresented in research.

The study had several limitations of note. First, a comprehensive psychometric validation of the administered measures was not conducted. Authors of prior work obtained 1-week retest reliability and administered a broader battery to assess convergent and divergent validity (Shapiro et al., 2013). Future work should attempt to replicate associations between psychopathology and other constructs that have been established in previous work and conduct further measure validation. Second, we were unable to independently verify if each participant experienced a PTE; thus, a portion of our sample may have provided inaccurate answers to receive compensation. However, this is a risk in many different domains of assessment, including phone and in-person interviews. To protect against this limitation, several metrics of validity were used, such as including questions that warranted a specific response, monitoring the time to completion, keeping the offered compensation offered while remaining consistent with other MTurk survey HITs, and using branching logic to remove participants who did not meet the inclusion criteria. Additional research is needed to devise a comprehensive set of strategies to improve the validity of the obtained responses. Experts in this area have begun to develop these approaches, but more work is needed to verify their success. The online, survey-based design limited the present study to the use of self-report screener items to identify prevalence rates of psychopathology. Screening measures, by design, tend to produce higher rates of psychopathology than do face-to-face clinical interviews. By contrast, the study design of Kilpatrick and colleagues (2013) was intended to closely match a clinical interview, which may account for the lower rates of PTSD they observed in their sample. The prevalence rates of psychopathology reported in the present study should thus be interpreted in light of this possible measurement bias. In addition, the study conducted by Santiago and colleagues (2013) examined different time frames (i.e., PTSD symptoms reported between 3 and 12 months posttrauma) than were used in the present study in order to generate a mean 12-month PTSD prevalence rate; these differences should be taken into consideration when comparing prevalence rates between studies. Another limitation is the possibility of self-selection bias, in that individuals may have been more likely to select into the HIT if they were highly confident that their traumatic exposure applied whereas those who were unsure if they would screen in may not have accessed the HIT at all. Finally, the present study was limited in that it did not include an explicit comparison of the prevalence rates of psychopathology found in the MTurk sample to epidemiological study designs versus convenience sample study designs. Future work should examine the degree to which MTurk samples compare to these approaches when employing strategies utilized by each (e.g., weighting the sample to be representative of the general population).

Taken together, the results of the present study suggest that crowdsourcing methods are a viable strategy to recruit a PTE-exposed sample and facilitate collection of samples with clinically meaningful symptoms. This approach is especially useful as a means to collect large quantities of data in a brief time period with a limited budget. Such a strategy is optimal for collecting preliminary data for novel hypotheses in preparation for costly research. As with any method, the strengths of crowdsourced data collection should be considered in light of its limitations. Researchers must take additional, crowdsourcing-specific steps to ensure the quality of their data. When data validation practices are used, however, high quality data from hard-to-reach samples can be collected.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by a grant awarded to Andrew Brown from the University of Vermont Academic Programs for Learning and Engagement Award. Katherine van Stolk-Cooke and Matthew Price were supported by 1K08MH107661-01A1. Rex Forehand was supported by R01MH100377, and Justin Parent was supported by NICHD R01HD064723. Anne Maheux was supported by the APLE Summer Stipend from the University of Vermont.

References

Armour C, Tsai J, Durham T, Charak R, Biehn T, Elhai J, Pietrzak R. Dimensional structure of *DSM-5* posttraumatic stress symptoms: Support for a hybrid anhedonia and externalizing behaviors model. Journal of Psychiatric Research. 2015; 61:106–113. DOI: 10.1016/j.jpsychires.2014.10.012 [PubMed: 25479765]

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th. Arlington, VA: Author; 2013.

Batterham PJ. Recruitment of mental health survey participants using Internet advertising: Content, characteristics and cost effectiveness. International Journal of Methods in Psychiatric Research. 2014; 23:184–191. DOI: 10.1002/mpr.1421 [PubMed: 24615785]

Berinsky A, Huber G, Lenz G. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. Political Analysis. 2012; 20:351–368. DOI: 10.1093/pan/mpr057

Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for US adults: national health interview survey, 2012. Vital and Health Statistics. Series 10, Data from the National Health Survey. 2014; 260:1–161.

Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The posttraumatic stress disorder checklist for *DSM-5* (PCL-5): Development and initial psychometric evaluation. Journal of Traumatic Stress. 2015; 28:489–498. DOI: 10.1002/jts.22059 [PubMed: 26606250]

Bovin MJ, Marx BP, Weathers FW, Gallagher MW, Rodriguez P, Schnurr PP, Keane TM. Psychometric properties of the PTSD Checklist for *Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition* (PCL-5) in veterans. Psychological Assessment. 2016; 28:1379–1391. DOI: 10.1037/pas0000254 [PubMed: 26653052]

Breslau N, Troost JP, Bohnert K, Luo Z. Influence of predispositions on post-traumatic stress disorder: does it vary by trauma severity? Psychological Medicine. 2013; 43:381–390. DOI: 10.1017/S0033291712001195 [PubMed: 22703614]

Bryant RA, McFarlane AC, Silove D, O'Donnell ML, Forbes D, Creamer M. The lingering impact of resolved PTSD on subsequent functioning. Clinical Psychological Science. 2016; 4:493–498. DOI: 10.1177/2167702615598756

Chandler J, Mueller P, Paolacci G. Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. Behavior Research Methods. 2014; 46:112–130. DOI: 10.3758/s13428-013-0365-7 [PubMed: 23835650]

Chandler J, Shapiro D. Conducting Clinical Research Using Crowdsourced Convenience Samples. Annual Review of Clinical Psychology. 2016; 12:53–81. DOI: 10.1146/annurev-clinpsy-021815-093623

Elhai JD, Palmieri PA. The factor structure of posttraumatic stress disorder: A literature update, critique of methodology, and agenda for future research. Journal of Anxiety Disorders. 2011; 25:849–854. DOI: 10.1016/j.janxdis.2011.04.007 [PubMed: 21793239]

Flora D, Curran P. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods. 2004; 9:466.doi: 10.1037/1082-989X.9.4.466 [PubMed: 15598100]

Friedman MJ, Resick PA, Bryant RA, Brewin CR. Considering PTSD for *DSM-5*. Depression and Anxiety. 2011; 28:750–769. DOI: 10.1002/da.20767 [PubMed: 21910184]

Garrett SKM, Thomas AP, Cicuttini F, Silagy C, Taylor HR, McNeil JJ. Community-based recruitment strategies for a longitudinal interventional study: The VECAT experience. Journal of Clinical Epidemiology. 2000; 53:541–548. DOI: 10.1016/S0895-4356(99)00153-5 [PubMed: 10812328]

Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 1999; 6:1–55. DOI: 10.1080/10705519909540118

Huff C, Tingley D. "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. Research & Politics. 2015; 2(3): 2053168015604648.doi: 10.1177/2053168015604648

Jones L, Veinot T, Pressler S, Seng J, McCall A, Fernandez D, Coleman-Burns P. Internet Health Information Seeking (IHIS) An Integrative Review of the Literature. Western Journal of Nursing Research. 2014; 36:1376–1377. DOI: 10.1177/0193945914540100 [PubMed: 25288618]

Kang SH, Gratch J. Virtual humans elicit socially anxious interactants' verbal self-disclosure. Computer Animation and Virtual Worlds. 2010; 21:473–482. DOI: 10.1002/cav.345

Kass RE, Raftery AE. Bayes factors. Journal of the American Statistical Association. 1995; 90:773–795. DOI: 10.1080/01621459.1995.10476572

Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the National Comorbidity Survey. Archives of General Psychiatry. 1995; 52:1048–1060. DOI: 10.1001/archpsyc.1995.03950240066012 [PubMed: 7492257]

Kilpatrick D, Resnick H, Milanak M, Miller M, Keyes K, Friedman M. National estimates of exposure to traumatic events and PTSD prevalence using *DSM-IV* and *DSM-5* criteria. Journal of Traumatic Stress. 2013; 26(5):537–547. DOI: 10.1002/jts.21848 [PubMed: 24151000]

King DW, Leskin GA, King LA, Weathers FW. Confirmatory factor analysis of the clinician-administered PTSD Scale: Evidence for the dimensionality of posttraumatic stress disorder. Psychological Assessment. 1998; 10:90–96.

Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, Clark LA. The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. Journal of Abnormal Psychology. 2017; 126:454–477. DOI: 10.1037/abn0000258 [PubMed: 28333488]

Kroenke K, Spitzer RL, Williams JB. The PHQ-9. Journal of General Internal Medicine. 2001; 16(9): 606–613. DOI: 10.1046/j.1525-1497.2001.016009606.x [PubMed: 11556941]

Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. Journal of Affective Disorders. 2009; 114:163–173. DOI: 10.1016/j.jad.2008.06.026 [PubMed: 18752852]

Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, Herzberg PY. Validation and Standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the General Population. Medical Care. 2008; 46:266–274. DOI: 10.1097/MLR.0b013e318160d093 [PubMed: 18388841]

Malcoun E, Houry D, Arndt-Jordan C, Kearns MC, Zimmerman L, Hammond-Susten M, Rothbaum BO. Feasibility of identifying eligible trauma patients for posttraumatic stress disorder

intervention. Western Journal of Emergency Medicine. 2010; 11(3):274–278. Retrieved from https://escholarship.org/uc/item/8pr618d4. [PubMed: 20882149]

Miller AK, Canales EJ, Amacker AM, Backstrom TL, Gidycz CA. Stigma-threat motivated nondisclosure of sexual assault and sexual revictimization: A prospective analysis. Psychology of Women Quarterly. 2011; 35:119–128. DOI: 10.1177/0361684310384104

Muthén LK, Muthén BO. Mplus Version 7 user's guide. Los Angeles, CA: Muthén & Muthén; 2012.

National Institutes of Mental Health. Posttraumatic stress disorder among adults. 2016. Retrieved from https://www.nimh.nih.gov/health/statistics/post-traumatic-stress-disorder-ptsd.shtml

Paolacci G, Chandler J, Ipeirotis PG. Running experiments on amazon Mechanical Turk. Judgment and Decision Making. 2010; 5:411–419.

Pietrzak RH, Goldstein RB, Southwick SM, Grant BF. Prevalence and Axis I comorbidity of full and partial posttraumatic stress disorder in the United States: Results from Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions. Journal of Anxiety Disorders. 2011; 25:456–465. DOI: 10.1016/j.janxdis.2010.11.010 [PubMed: 21168991]

Pietrzak RH, Tsai J, Armour C, Mota N, Harpaz-Rotem I, Southwick SM. Functional significance of a novel 7-factor model of *DSM-5* PTSD symptoms: Results from the National Health and Resilience in Veterans Study. Journal of Affective Disorders. 2015; 174:522–526. DOI: 10.1016/j.jad.2014.12.007 [PubMed: 25556669]

Price M, Ruggiero KJ, Ferguson PL, Patel SK, Treiber F, Couillard D, Fahkry SM. A feasibility pilot study on the use of text messages to track PTSD symptoms after a traumatic injury. General Hospital Psychiatry. 2014; 36:249–254. DOI: 10.1016/j.genhosppsych.2014.02.004 [PubMed: 24636721]

Price M, van Stolk-Cooke K. Examination of the interrelations between the factors of PTSD, major depression, and generalized anxiety disorder in a heterogeneous trauma-exposed sample using *DSM-5* criteria. Journal of Affective Disorders. 2015; 186:149–155. DOI: 10.1016/j.jad.2015.06.012 [PubMed: 26241663]

Price M, van Stolk-Cooke K, Ward HL, O'Keefe M, Gratton J, Skalka C, Freeman K. Tracking post-trauma psychopathology using mobile applications: A usability study. Journal of Technology in Behavioral Science. 2017; 2(41):1–8. DOI: 10.1007/s41347-016-0008-9

Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. Journal of Medical Internet Research. 2012; 14:e28.doi: 10.2196/jmir.1878 [PubMed: 22360969]

Roberts AL, Gilman SE, Breslau J, Breslau N, Koenen KC. Race/ethnic differences in exposure to traumatic events, development of post-traumatic stress disorder, and treatment-seeking for post-traumatic stress disorder in the United States. Psychological Medicine. 2011; 41:71–83. DOI: 10.1017/S0033291710000401 [PubMed: 20346193]

Rothbaum BO, Kearns MC, Price M, Malcoun E, Davis M, Ressler KJ, Houry D. Early intervention may prevent the development of posttraumatic stress disorder: a randomized pilot civilian study with modified prolonged exposure. Biological Psychiatry. 2012; 72:957–963. DOI: 10.1016/j.biopsych.2012.06.002 [PubMed: 22766415]

Santiago PN, Ursano RJ, Gray CL, Pynoos RS, Spiegel D, Lewis-Fernandez R, Fullerton CS. A systematic review of PTSD prevalence and trajectories in *DSM-5* defined trauma exposed populations: intentional and non-intentional traumatic events. PloS One. 2013; 8(4):e59236.doi: 10.1371/journal.pone.0059236 [PubMed: 23593134]

Shapiro DN, Chandler J, Mueller PA. Using Mechanical Turk to study clinical populations. Clinical Psychological Science. 2013; 1:213–220. DOI: 10.1177/2167702612469015

Skinner HA, Allen BA. Does the computer make a difference? Computerized versus face-to-face versus self-report assessment of alcohol, drug, and tobacco use. Journal of Consulting and Clinical Psychology. 1983; 51:267–275. DOI: 10.1037/0022-006X.51.2.267 [PubMed: 6841771]

Spitzer RL, Kroenke K, Williams JW, Lowe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine. 2006; 166:1092–1097. DOI: 10.1001/archinte.166.10.1092 [PubMed: 16717171]

Pew Research Center. Internet/broadband fact sheet. 2017a. Retrieved from http://www.pewinternet.org/fact-sheet/mobile

Pew Research Center. Mobile fact sheet. 2017b. Retrieved from http://www.pewinternet.org/fact-sheet/mobile/

Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. Science. 1998; 280:867–873. DOI: 10.1126/science.280.5365.867 [PubMed: 9572724]

Walsh K, Danielson CK, McCauley JL, Saunders BE, Kilpatrick DG, Resnick HS. National prevalence of posttraumatic stress disorder among sexually revictimized adolescent, college, and adult household-residing women. Archives of General Psychiatry. 2012; 69:935–942. DOI: 10.1001/archgenpsychiatry.2012.132 [PubMed: 22566561]

Weathers FW, Blake DD, Schnurr PP, Kaloupek DG, Marx BP, Keane TM. The life events checklist for DSM-5 (LEC-5). 2013. Available from the National Center for PTSD website at http://www.ptsd.va.gov

Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD Checklist for DSM-5 (PCL-5). 2013. Available from the National Center for PTSD website at http://www.ptsd.va.gov

Yufik T, Simms LJ. A meta-analytic investigation of the structure of posttraumatic stress disorder symptoms. Journal of Abnormal Psychology. 2010; 119:764–776. DOI: 10.1037/a0020981 [PubMed: 21090877]

Table 1

Demographic Statistics and Trauma Exposure

Variable	*n*	%
Men	371	45.1
Race/Ethnicity		
White, Non-Hispanic	607	73.8
Hispanic/Latino, White	50	6.1
Hispanic/Latino, African American	7	0.9
African American, Non-Hispanic	57	6.9
Asian	69	8.4
American Indian or Native Alaskan	12	1.5
Other	18	2.2
Relationship Status		
Single	333	40.5
Married	296	36.0
Cohabitating for > 1 year	125	15.2
Divorced	45	5.5
Separated	7	0.9
Widowed	5	0.6
Employment status		
Full-time employed	485	59.0
Part-time employed	95	11.6
Full-time student	48	5.8
Part-time student	7	0.9
Homemaker	55	6.7
Unemployed	80	9.7
Disabled	11	1.3
Retired	8	1.0
Education		
High school diploma	93	11.3
1–2 years of college	270	32.8
3–4 years of college	327	39.8
Some graduate school	43	5.2
Master’s degree	66	8.0
Advanced degree	14	1.7
Income (USD)		
< $25,000	215	26.2
$25,000–50,000	293	35.6
$50,000–75,000	163	19.8
$75,000–100,000	81	9.9
> $100,000	65	7.9
Reported a history of treatment for PTSD	80	9.7

Variable	*n*	%
Age, years (*M* and *SD*)	34.28	10.79

Note. PTSD = posttraumatic stress disorder.

Table 2

Frequencies of Trauma Exposure Reported in the Sample

Trauma Exposure	Experienced		Witnessed		Learned		Part of Job		Not Sure		N/A	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Natural disaster	324	39.4	132	16.1	139	16.1	10	1.2	9	1.1	187	22.7
Fire or explosion	112	13.6	185	22.5	154	18.7	10	1.2	9	0.9	314	38.2
Vehicle accident	502	61.1	127	15.5	63	7.7	9	1.1	3	0.4	99	12.0
Accident (other)	157	19.1	145	17.6	145	17.6	10	1.2	22	2.7	305	37.1
Toxic substance	43	5.2	20	2.4	125	15.2	34	4.1	27	3.3	516	62.8
Assault	311	37.8	106	12.9	123	15.0	9	1.1	11	1.3	227	27.6
Assault (weapon)	117	14.2	62	7.5	152	18.5	10	1.2	9	1.1	421	51.2
Sexual Assault	159	19.3	31	3.8	186	22.6	10	1.2	8	1.0	384	46.7
Unwanted sexual experience	257	31.3	28	3.4	113	13.7	13	1.6	11	1.3	351	42.7
Combat/war zone	18	2.2	18	2.2	175	21.3	12	1.5	8	1.0	520	63.3
Captivity	21	2.6	8	1.0	109	13.3	5	0.6	13	1.6	593	72.1
Illness/injury	119	14.5	206	25.1	127	15.5	13	1.6	3	0.4	301	36.6
Severe suffering	38	4.6	98	11.9	130	15.8	15	1.8	19	2.3	449	54.6
Death (violent)	21	2.6	101	12.3	216	26.3	10	1.2	5	0.6	409	49.8
Death (accidental)	30	3.6	109	13.3	214	26.0	14	1.7	6	0.7	387	47.1
Caused harm	30	3.6	44	5.4	77	9.4	10	1.2	4	0.5	578	70.3
Other	304	37.0	53	6.4	47	5.7	12	1.5	41	5.0	311	37.8

Table 3

Factor Loadings for Posttraumatic Stress Disorder (PTSD) Symptoms

Symptom	*DSM-5* Model	*DSM-5* Cluster	Hybrid Model	Hybrid Cluster
Intrusive thoughts	.860	Intrusions	.861	Intrusions
Nightmares	.820	Intrusions	.821	Intrusions
Reliving trauma	.869	Intrusions	.869	Intrusions
Emotional cue reactivity	.878	Intrusions	.878	Intrusions
Physiological cue reactivity	.856	Intrusions	.856	Intrusions
Avoidance (internal cues)	.877	Avoidance	.878	Avoidance
Avoidance (external cues)	.944	Avoidance	.944	Avoidance
Trauma-related amnesia	.637	NACM	.651	Negative affect
Negative view of self/world	.833	NACM	.856	Negative affect
Blame self/others	.692	NACM	.706	Negative affect
Pervasive negative emotions	.861	NACM	.880	Negative affect
Loss of interest	.889	NACM	.912	Anhedonia
Detachment	.879	NACM	.896	Anhedonia
Anhedonia	.907	NACM	.927	Anhedonia
Irritability/anger	.885	AAR	.898	Externalizing
Recklessness	.733	AAR	.750	Externalizing
Overly alert	.744	AAR	.815	Anxious arousal
Easily startled	.826	AAR	.924	Anxious arousal
Difficulty concentrating	.856	AAR	.913	Dysphoric arousal
Sleep problems	.794	AAR	.843	Dysphoric arousal
BIC	37822.235		37507.279	

Note. BIC = Bayesian information criterion.