



Published in final edited form as:

Int J Drug Policy. 2020 January ; 75: 102587. doi:10.1016/j.drugpo.2019.10.013.

Leveraging Crowdsourcing Methods to Collect Qualitative Data in Addiction Science: Narratives of Non-Medical Prescription Opioid, Heroin, and Fentanyl Use

Justin C. Strickland^{1,*}, Grant A. Victor²

¹Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, 5510 Nathan Shock Drive, Baltimore, MD 21224, USA

²Center for Behavioral Health and Justice, Wayne State University School of Social Work, 5201 Cass Ave, Detroit, Michigan, 48202, USA

Abstract

Background: Online crowdsourcing methods have proved useful for studies of diverse designs in the behavioral and addiction sciences. The remote and online setting of crowdsourcing research may provide easier access to unique participant populations and improved comfort for these participants in sharing sensitive health or behavioral information. To date, few studies have evaluated the use of qualitative research methods on crowdsourcing platforms and even fewer have evaluated the quality of data gathered. The purpose of the present analysis was to document the feasibility and validity of using crowdsourcing techniques for collecting qualitative data among people who use drugs.

Methods: Participants (N = 60) with a history of non-medical prescription opioid use with transition to heroin or fentanyl use were recruited using Amazon Mechanical Turk (mTurk). A battery of qualitative questions was included indexing beliefs and behaviors surrounding opioid use, transition pathways to heroin and/or fentanyl use, and drug-related contacts with structural institutions (e.g., health care, criminal justice).

Results: Qualitative data recruitment was feasible as evidenced by the rapid sampling of a relatively large number of participants from diverse geographic regions. Computerized text analysis indicated high ratings of authenticity for the provided narratives. These authenticity percentiles were higher than the average of general normative writing samples as well as than those collected in experimental settings.

*To whom correspondence should be addressed: Justin C. Strickland, Behavioral Pharmacology Research Unit, Johns Hopkins University School of Medicine, 5510 Nathan Shock Drive Baltimore, MD 21224-6823, USA. Phone: (410) 550-1975
jstric14@jhmi.edu.

Contributors

J.C. Strickland and G.A. Victor developed the study concept. Data were collected by J.C. Strickland. J.C. Strickland conducted the primary data analysis with additional guidance provided by G.A. Victor. J.C. Strickland drafted the initial manuscript with critical feedback from G.A. Victor. All authors approved the final version of the manuscript for submission.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declarations of Interest: None

Conclusions: These findings support the feasibility and quality of qualitative data collected in online settings, broadly, and crowdsourced settings, specifically. Future work among people who use drugs may leverage crowdsourcing methods and the access to hard-to-sample populations to complement existing studies in the human laboratory and clinic as well as those using other digital technology methods.

Keywords

Crowdsourcing; Heroin; Mechanical Turk; Opioid; Qualitative; Mixed Method

Introduction

The past decade has witnessed an exponential growth in the use of Internet crowdsourcing tools in behavioral and biomedical research. Broadly defined, crowdsourcing refers to the use of an open call to large groups of people in order to complete an otherwise difficult to accomplish task (Estellés-Arolas & González-Ladrón-De-Guevara, 2012). Researchers have recognized the benefits that crowdsourcing may hold when used as a recruitment tool for identifying and sampling research participants. In a typical example of crowdsourced sampling, a researcher posts a research study on a crowdsourcing platform (e.g., Amazon Mechanical Turk [mTurk], Facebook, Prolific) made visible to any potential participant that is a member of that platform. This process closely mirrors the recruitment strategies typically used in Psychology 101 pools, but affords a sampling space with participants with greater variation in location, demographics, and health histories within an integrated online platform (Casler, Bickel, & Hackett, 2013; Landers & Behrend, 2015; Smith, Sabat, Martinez, Weaver, & Xu, 2015). As such, crowdsourcing has proved useful for a wide range of disciplines from addiction science and clinical psychological to education research and cancer biology (see reviews by Chandler & Shapiro, 2016; Follmer, Sperling, & Suen, 2017; Y. J. Lee, Arida, & Donovan, 2017; Miller, Crowe, Weiss, Maples-Keller, & Lynam, 2017; Strickland & Stoops, 2019).

One of the most popular of these sources for addiction science research is Amazon Mechanical Turk (mTurk) (Chandler & Shapiro, 2016; Strickland & Stoops, 2019). mTurk is a platform developed by Amazon as an online labor market allowing individuals and businesses to outsource various problems to a national and international human workforce. Researchers have leveraged this online setting to post research opportunities that are completed completely through the Internet interface. Such research is incentivized through monetary compensation for participation that is handled completely within the mTurk setting.

The ease and efficiency of online sampling in crowdsourcing, broadly, and on mTurk, specifically, means that large sample sizes may be generated rapidly thereby making crowdsourcing a popular tool for measure development research (e.g., exploratory or confirmatory factor analyses; Dunn, Barrett, Herrmann, et al., 2016; Dunn, Barrett, Yopez-Laubach, et al., 2016; Gearhardt, Corbin, & Brownell, 2016). The use of open-source, browser-based programming tools such as PsyToolKit allows for the evaluation of reaction time data and application of cognitive-behavioral tasks through these Internet platforms

(e.g., Crump, McDonnell, & Gureckis, 2013; Stewart, Chandler, & Paolacci, 2017; Stoet, 2017; Strickland, Bolin, Lile, Rush, & Stoops, 2016). Although the majority of studies using mTurk and crowdsourcing remain cross-sectional in design, recent work has demonstrated the feasibility and validity of collecting naturally occurring intensive longitudinal data (e.g., Hartsell & Neupert, 2017; Lanaj, Johnson, & Barnes, 2014; Strickland & Stoops, 2018b) as well as data following behavioral interventions targeting substance use disorder (e.g., Cunningham, Godinho, & Bertholet, 2019; Cunningham, Godinho, & Kushnir, 2017; Strickland, Hill, Stoops, & Rush, 2019). Importantly, prior research has also demonstrated a close correspondence between the results of in-person research and online research on mTurk among people who use drugs (e.g., Jarmolowicz, Bickel, Carter, Franck, & Mueller, 2012; Patrick S Johnson, Herrmann, & Johnson, 2015; Kim & Hodgins, 2017; Morris et al., 2017; Strickland et al., 2016; Strickland, Lile, & Stoops, 2017).

Despite this proliferation of crowdsourced research in addiction science and applications across varied methodologies, remarkably little research has been conducted using qualitative or mixed method techniques. This is unfortunate given the benefits of qualitative research for understanding individuals' lived experiences relevant to substance use disorder (Rhodes, Stimson, Moore, & Bourgois, 2010) and the specific benefits that crowdsourcing may offer to this end. Crowdsourcing allows improved access to hard-to-recruit populations providing for more rapid recruitment of relevant and adequately sized samples at low monetary and time costs. Recruitment may also proceed without reliance on snowball sampling thereby avoiding problems with bias or homogenous sample characteristics those techniques can produce (Biernacki & Waldorf, 1981; Faugier & Sargeant, 1997). More broadly, computerized and remote delivery reduces burden for potential participants, particularly those who may lack the time for in-person assessments or experience barriers to reaching the research site due to financial, cultural, or health reasons (Areán & Gallagher-Thompson, 1996; Blanton et al., 2006). The confidentiality¹ of online platforms and computerized data collection may also allow for increased trust and comfort when sharing sensitive information such as sexual, psychiatric, or substance use histories (Harrison & Hughes, 1997; Turner et al., 1998).

Preliminary applications of qualitative research on crowdsourcing platforms show good promise for utilizing these methods. Existing studies have shown feasibility in targeting unique populations such as survivors of childhood sexual abuse (Schnur, Dillon, Goldsmith, & Montgomery, 2017). However, only one study has formally evaluated and compared the quality of data collected using qualitative methods across online and in-person settings (Grysmann, 2015). Participants in that study wrote narratives of a stressful life event and content was evaluated between mTurk and college-student sources. mTurk participants wrote shorter narratives but reported events that were more stressful and that used more negative emotion language than the college student samples. These findings are consistent with the idea that mTurk is feasible and provides a valid and potentially unique opportunity to collect qualitative data easily and efficiently. The generality of this finding remains unknown for other populations, especially those reporting sensitive health behaviors like substance use.

¹Data from most crowdsourcing resources is not strictly anonymous given the use of ID names or numbers (e.g., mTurk Worker IDs).

The purpose of the present manuscript was to document the feasibility and validity of utilizing crowdsourcing methods for collecting qualitative data among people who use drugs. The specific aims of the conducted analyses included: 1) presenting the feasibility of screening and general procedures involved in collection of qualitative data using crowdsourcing methods; 2) assessing the quality of the qualitative data through computerized text analysis; and 3) evaluating the implications for and utility of qualitative crowdsourced data in the field of addiction science. Although we hypothesized that crowdsourcing would be useful for qualitative research across broad substance use contexts, this study focused on one example of illicit substance use that carries current and salient public health implications. This substantive research context was a project designed to augment current understanding of the overdose crisis in the United States. To this end, participants who reported a history of non-medical prescription opioid use (NMPOU) and a subsequent transition to heroin and/or fentanyl use were sampled using crowdsourcing methods. Shifts from NMPOU to the use of heroin and synthetic opioids like fentanyl have been closely implicated in national increases in overdose fatalities in the United States; however, relatively little work exists understanding the qualitative and narrative experience of these transitional pathways. A series of open-ended questions were used to collect accounts of substance use circumstances, reasoning, social and legal/policy contexts, and harm reduction approaches relevant to opioid use. We predicted that crowdsourcing methods would be both feasible and provide useful and valid data based on prior successes of applying crowdsourcing in addiction science.

Methods

Screening and General Procedures

Participants were recruited from mTurk between 4 March 2019 and 1 May 2019. Eligibility to view the study was limited to individuals from the United States who had completed 100 or more prior mTurk tasks with a 99% or greater approval rating on those tasks. These restrictions involving geographic location and task completion are implemented directly through the mTurk portal. Automated responding was limited by use of a CAPTCHA response requirement to advance the initial survey page. Sampling was limited to the United States given the parent pilot study's focus on patterns of substance use observed at that national level (i.e., narratives of transitions from NMPO to heroin/fentanyl use). However, mTurk and many other crowdsourcing sources are not limited to a United States population and may be used to generate international samples through open recruitment or country-specific targeting. Completion and approval criteria were used to improve quality and attention and were used consistent with prior uses and recommendations on mTurk (e.g., Kaplan et al., 2017; Morean, Lipshie, Josephson, & Foster, 2017; Peer, Vosgerau, & Acquisti, 2014; Strickland & Stoops, 2019). These inclusion criteria are typically used in crowdsourced research to improve data quality as well as prevent automated responding (e.g., "bots"). Specifically, prior research has found that individuals with higher task completion or approval rating had higher reliability scores for previously developed measures, failed fewer attention checks, and showed lower rates of "problematic" responding (e.g., central tendency bias) (Peer et al., 2014). Other research evaluating samples collected with and without these criteria has identified few differences on

demographic or substance use measures (Peer et al., 2014; Strickland & Stoops, 2018a) suggesting that bias is likely limited. Participants that met the task and approval criteria were able to see the posted study description on the mTurk task database. The study description was titled “Bonus Available. Answer questions and complete tasks about behavior” with a summary depicting a general survey about feelings, attitudes, and behaviors (see Supplemental Materials). No overt eligibility criteria concerning substance use history were included in this description. No messages or direct contact was made with participants to initiate participation.

Participants then completed a screening questionnaire that included questions about substance use history as well as other health behaviors like dietary and sleep habits (to further mask eligibility criteria). The inclusion criteria for this study were 1) age 18 or older, 2) lifetime non-medical prescription opioid use, and 3) transition to heroin and/or fentanyl use following non-medical prescription opioid use initiation (95% of participants reported heroin use). Gender identity was not an inclusion/exclusion criterion, although only male and female participants were represented in the qualifying sample. Qualifying participants were directed to the full study survey containing qualitative opioid use history questions and a general health history. All participants were compensated \$0.05 for completing the study screener and qualifying participants were compensated \$5 for completing the full study. The University of Kentucky Institutional Review Board reviewed and approved all study procedures.

Qualitative Questions: Opioid Use History

Participants completed a series of nine semi-structured, open-ended questions regarding substance use history. These questions were designed to record beliefs and behaviors relevant to NMPOU and transitions to heroin and/or fentanyl use. Specific items asked for information such as initiation of use, administration routes, drug use relationships, and contact with drug-related criminal justice system, health care setting(s) and drug treatment services (see Supplemental Materials for instructions and full question set). Instructions stipulated that there were no length requirements, but to be as thorough as possible. Participants were also told that they did not have to answer a question if it did not apply (e.g., if they had no contact with the criminal justice system). As noted in the Introduction, analyses and data presented here focused on the feasibility and quality of this qualitative data collection. Results of qualitative coding for emergent themes regarding substance use will be reported in a forthcoming manuscript.

Computerized Text Analysis

Computerized text analysis was conducted using Linguistic Inquiry Word Count (LIWC) 2015 software (Pennebaker, Boyd, Jordan, & Blackburn, 2015; Tausczik & Pennebaker, 2010). LIWC was selected given its extensive validation testing over the past decades and widespread use in the field of textual analysis. Qualitative responses were combined from all nine questions to improve precision of proportional word count estimates by increasing the base word counts. Overt spelling errors identified using Microsoft Office spellcheck were corrected prior to analysis (e.g., opiod to opioid). Outputs of LIWC analyses using the default dictionary included linguistic processes (total word count, words/sentence, first-

person singular language), psychological processes (% affective language [total, positive, and negative]), and authenticity percentiles. Linguistic and psychological processes were selected prior to analysis as those broadly relevant to personal narratives and data acquisition for these qualitative histories.

Authenticity scores are provided based on an algorithm derived from prior empirical work on deception and honesty in written and spoken language (Newman, Pennebaker, Berry, & Richards, 2003). These scores quantify the degree to which participants are believed to respond in an authentic or honest manner. Scores are percentiled with higher values representing higher expected authenticity.

Data Analysis

Data from participants completing the full survey ($N = 66$) were first examined for attentive and systematic data using attention and validity checks. These included: 1) comparison of age and gender identity at two separate locations in the survey, 2) consistency in self-reported history of NMPOU, and 3) non-response (i.e., blank responding) or illogical responses on qualitative questions (e.g., the single word “NICE” as a response to all questions). Six participants (9.1%) were identified as inattentive or non-systematic and removed from subsequent analysis.

Demographic and substance use histories from systematic participants ($N = 60$) were first evaluated using descriptive statistics. Database values were then gathered from the LIWC development and psychometric properties manual for comparison to values generated in the current study. These database estimates are based on 117,779 files containing a total of 231,190,022 words from sources including blogs, expressive writing, novels, natural speech, *New York Times*, and Twitter (Pennebaker et al., 2015). Comparisons were made with the grand mean from these database values as well as specifically with those from expressive writing samples. Expressive writing data were considered particularly relevant because these data were taken from experimental studies in which participants were asked to write about deeply emotional topics (e.g., a personally upsetting experience) (6179 total files analyzed in the LIWC database).

Analyses compared 95% confidence intervals to LIWC database values with significance tests conducted as one-sample *t*-tests and effects sizes calculated as Cohen's *d*. Additional analyses compared authenticity percentiles as a primary outcome variable to the LIWC database for all text source types. Analyses were conducted using LIWC 2015 (Pennebaker et al., 2015) and *R* statistical language.

Results

Participants completed the study survey in an average of 31.3 minutes ($SD = 19.6$). Characteristics of recruited participants are presented in Table 1. A majority of participants were White and reported smoking tobacco cigarettes. Approximately half were female, and half completed a high school education or less. These departures from general population demographics, such as higher rates of tobacco cigarette use and lower education rates, are consistent with expected characteristics of individuals with a history of NMPOU.

Participants self-reported being from 27 different states with Florida ($n = 6$) as the most popular followed by Pennsylvania ($n = 5$) (see Supplemental Materials for complete distribution).

Approximately half of participants reported current NMPOU and half reported current heroin use. A majority reported lifetime heroin use (95%), injection drug use (60%), and overdose (54.2%). One-third reported current medication-assisted treatment (methadone, buprenorphine, or naltrexone).

Computerized Text Analysis

Participants wrote, on average, 377 words across the nine questions, although variability was observed between participants and between questions ($SD = 308.6$; Range = 14 to 1341 words total; see Supplemental Figure 1). Quantitative results from the computerized text analysis and LIWC database values are presented in Table 2. Compared to database values the mTurk sample used a greater percentage of first-person singular language and lower percentage of affective language, generally, and positive affective language, specifically, p values $< .001$, d values = 0.65–1.49. A similar pattern of results was observed when comparing to LIWC database values for expressive writing, although these differences were of a smaller magnitude effect size, p values $< .05$, d values = 0.27–0.50. mTurk writing samples did not differ on negative affective language with either the general LIWC database or expressive writing values.

Figure 1 contains individual authenticity percentiles from the mTurk sample as well as LIWC database comparisons for all writing types available. Visual inspection of these values indicated a high proportion of mTurk samples above the 80th percentile with a negative skew in the distribution (i.e., clustering in the upper range with a few low percentile values pulling central tendency downwards). Comparisons to all LIWC database sources indicated higher authenticity percentile scores in the mTurk sample at a medium-to-large effect size, p values $< .001$, d values > 0.50 .

Discussion

The purpose of the present study was to evaluate the feasibility and validity of using crowdsourced sampling to collect qualitative data among people who use drugs. Participants with a history of opioid use were feasibly recruited on the popular crowdsourcing platform mTurk. Computerized text analysis of qualitative data further highlighted the authenticity of provided narratives as well as specific linguistic patterns relevant to substance use disorder and drug-related lived experiences. These findings collectively contribute to and advance a growing literature demonstrating the flexibility of crowdsourcing platforms in behavioral research by establishing their viability for qualitative research, broadly, and conducting that research with difficult-to-sample populations in addiction science, specifically.

Recruitment of individuals who use(d) NMPOU was feasible on the crowdsourcing platform mTurk. This feasibility was established through the sampling of a relatively large number of participants necessary to achieve ‘Open coding’ (Corbin & Strauss, 1990) and data saturation (Fusch & Ness, 2015; Saunders et al., 2018) as well as by sampling from diverse

geographic regions with a low proportion of these individuals providing non-systematic or inattentive data. This finding is consistent with the extant crowdsourcing literature, which has described the benefits of crowdsourcing platforms for identifying and recruiting numerous specialized participant populations (e.g., cancer survivors or fathers as primary caregivers; Arch & Carr, 2017; Parent, Forehand, Pomerantz, Peisch, & Seehuus, 2017).

Research in addiction science specifically has used mTurk to generate samples with widely varying substance use and behavioral histories relevant to substance use disorder (Amlung et al., 2019; Bergeria, Huhn, & Dunn, 2019; Huhn, Garcia-Romeu, & Dunn, 2018; P. S. Johnson & Johnson, 2014; Morris et al., 2017; Strickland & Stoops, 2015). Qualitative addiction research demography is often skewed toward male participants and to balance sampling, efforts are made to over-recruit certain population groups, such as women (e.g., Mars, Bourgois, Karandinos, Montero, & Ciccarone, 2014). The slight majority of the current sample identified as female (56.7%) which indicates that this crowdsourcing tool may be preferable to traditional methods in recruiting female-balanced samples. Although not formally investigated here, prior research has also demonstrated the feasibility of using mTurk to sample individuals from sexual and gender minority groups (e.g., Andersen, Zou, & Blossnich, 2015; Catalpa et al., 2019; Rainey, Furman, & Gearhardt, 2018). The use of crowdsourcing resources also provides a unique opportunity to interface with individuals outside typical research or clinical settings. For example, individuals may be reached prior to a decision to engage in treatment within a clinical context. The rapid time of recruitment combined with demographic variability of this sample, including variations in recovery progression and behavioral history (e.g., injection drug use and overdose experience), highlights this capacity to more efficiently recruit populations with varying histories of interest through this sampling mechanism. As noted in the Introduction, crowdsourcing provides this clear benefit of allowing for much larger samples in a quicker manner and at lower cost than community-based approaches typically afford.

Participants provided narratives that were rated high in authenticity according to a computerized textual analysis and multivariate linguistic algorithm of deceptive responding. This finding was observed at both the group level with an average authenticity score in the 84th percentile as well as at the individual participant level with 80% of participants in the upper quartile of scores (75th percentile or greater). These high levels of expected authenticity held when relating to a variety of writing types from the LIWC database. Comparisons indicating higher ratings relative to expressive writing samples were especially encouraging given that these expressive writing samples were collected in a manner analogous to the methods used to collect qualitative data for this study (e.g., emotional writing collected in an experimental setting).

The findings regarding the expected authenticity may be of particular importance to qualitative researchers who are cautious to utilize crowdsourcing due to legitimate concerns about the length of immersion at research sites. That is, the current findings indicate that high levels of authenticity in the reporting of lived experiences can be achieved in the absence of immersion of research personnel. Therefore, the methodologies presented in the current study may be useful to researchers interested in recruiting participants from various communities, especially when travel and direct immersion is not feasible.

Highlighting the authenticity of qualitative responding is also relevant not just for qualitative research as described here, but for research conducted on mTurk generally. This is because problems related to deceptive responding remain salient and a primary concern when conducting crowdsourcing work. Biological or other objective verification of substance use and other health histories are not generally possible in an online setting. Therefore, studies must rely on participant self-report to verify inclusion/exclusion criteria and other study responses. On the one hand, the use of a remote, computerized delivery may improve the comfort that participants feel in responding truthfully. Research on mTurk has documented this comfort with several studies reporting that the majority of participants indicate they feel more comfortable sharing information of a sensitive nature online compared to in-person (Kim & Hodgins, 2017; Strickland et al., 2019; Strickland & Stoops, 2018b).

On the other hand, the incentive structure of mTurk and other crowdsourcing resources results in a situation where misrepresentation to meet eligibility criteria is monetarily reinforced. Several researchers have found that when these inclusion/exclusion criteria are made explicitly known, participants are more likely to engage in deception to gain study access (Chandler & Paolacci, 2017; Hydock, 2018; Sharpe Wessling, Huber, & Netzer, 2017). Mitigating these concerns are the use of best practice techniques, many of which we employed in this study. For example, one study found that while up to 89% of respondents misrepresented themselves when provided with the inclusion criteria needed to gain study access, less than 5% provided contradictory responses indicative of deception when no explicit criteria were available (Sharpe Wessling et al., 2017). It is relevant to note that evidence of deception under incentivized conditions is not unique to online settings and similar results have been observed in laboratory and field research (Fischbacher & Föllmi-Heusi, 2013; Mazar, Amir, & Ariely, 2008; Pruckner & Sausgruber, 2013). Collectively, the current study helps support this ongoing work by suggesting that when appropriate procedures – for example, concealing specific eligibility requirements and using a two-stage screening process – are used that honest and attentive quantitative and qualitative data can be generated in crowdsourcing settings.

Additional analyses examined other aspects of the linguistic patterns provided by participants. These findings indicated that qualitative responses were characterized by a lower proportion of affective language relative to the LIWC database estimates. This difference seemed mostly driven by lower proportions of positive affective language, whereas negative affective language was similar to general and expressive writing samples. Lower proportions of positively valenced affective language are consistent with the widely documented observation of anhedonia comorbid with NMPOU and substance use disorder patients (Garfield et al., 2017; Huhn et al., 2016; Sussman & Leventhal, 2014). An additional empirical study also suggests that affective language in written narratives is sensitive to substance use history and the recovery process. That study evaluated narratives written about stressful experiences as a part of a larger trial on mindfulness interventions for substance use disorder (Liehr et al., 2010). Use of positive affective language increased and negative affect language decreased among participants receiving the mindfulness training active group as well as in a historical control group that received a standard of care treatment (i.e., a therapeutic community intervention). The current findings related to affective language, though preliminary in nature, are largely in line with this prior work and provide

one example of how future qualitative work may leverage online sources for empirical purposes.

We utilized mTurk given its growing popularity in addiction science research (see review by Strickland & Stoops, 2019). However, it is important to consider the wealth of other digital technologies available for conducting qualitative research in the addiction sciences, and how mTurk may compare to these existing and already utilized web-based resources (see broad discussion of some web-based approaches in Barratt & Maddox, 2016; Coomber, 2011). One of such methods can be considered a “big data” technique which involves the extraction of existing information (i.e., web-scraping) available through public postings on social media websites like Twitter (e.g., Lamy et al., 2018; Sidani et al., 2019) or through general or specialty forums like Reddit (e.g., Costello, Martin III, & Brinegar, 2017; D’Agostino et al., 2017) and [Drugs-Forum.com](#) (e.g., Paul, Chisolm, Johnson, Vandrey, & Dredze, 2016). Web-scraping and related passive forms of data collection allow for the generation of large data sets that can be mined for relevant information such as behavioral patterns and narratives, albeit at the expense of the ability to ask direct or focused questions that is afforded through direct participant contact. To the latter point, there is also promising examples of digital methodology that involves the direct recruitment of individuals from online chat forums (e.g., Garcia-Romeu et al., 2019) or social media platforms like Facebook (e.g., Borodovsky, Marsch, & Budney, 2018; D. C. Lee, Crosier, Borodovsky, Sargent, & Budney, 2016; Ramo & Prochaska, 2012; Ramo et al., 2018). Also relevant to this body of research are studies involving direct interactions with participants through chat messaging allowing for dynamic interactions with individuals through text-based interviews (e.g., Barratt, 2012).

All of these digital sources, including mTurk, provide flexibility for browser-based data collection meaning that a variety of data types may be collected, including the potential for audio (Lane, Waibel, Eck, & Rottmann, 2010) or visual (Tran, Cabral, Patel, & Cusack, 2017) recordings (although the ethics/regulatory challenges posed by these kinds of data must be considered). mTurk is more limited than some of these digital technologies with respect to the sampling space (e.g., many more individuals are enrolled on Facebook than on mTurk). However, mTurk also benefits in that it was intentionally designed for sampling and survey/task completion purposes rather than designed for alternative purposes (e.g., social media) and utilized in unconventional ways for research (e.g., advertisement placements on Facebook for recruitment). One of the benefits of mTurk in this regard is the integrated nature of the platform that allows for recruitment and payment of participants within a single, secure setting. Use of other resources requires added complexity for payments or, often, recruitment of volunteers who will complete studies for free, which may impact generalizability of the sample. In this way, mTurk is similar to market research panels or panel data sets historically used in other social sciences like economics. Using mTurk as compared to specialty drug forums also allows for recruitment of a broader population of persons who use drugs than those who elect to participate in online forum communities that may be potentially biased by niche use patterns. Although other sampling bias related to who may choose to participate on mTurk should be considered (see more discussion of this issue below), mTurk may complement existing digital technologies by providing an integrated digital environment involving a behavioral diverse sampling space.

Limited work has also evaluated the quality of qualitative data generated using this collection of digital methodologies and much of this work was conducted during early utilizations of these resources (e.g., in the early 2000s). One study in marketing research, for example, found that as compared to mail or phone approaches, responses from web-based qualitative surveys were similar in complexity and richness of data (Coderre, Mathieu, & St-Laurent, 2004). We are not aware of research in addiction science that has similarly quantified the quality or richness of qualitative data collected through digital mediums. Some work, however, has found that qualitative interviews collected through web-based platforms among persons who use drugs and/or persons in recovery seem to provide meaningful information about substance use histories with participants viewing the online experience in a positive manner (e.g., Barratt, 2012; Dugdale, Elison, Davies, Ward, & Jones 2016). Therefore, the current study extends this work with digital technologies by providing a novel quantifiable indication of the quality of qualitative data about substance use histories as well as a possible framework for evaluating such richness in other digital contexts.

Another issue to consider in crowdsourced work is the payment structure. Little consensus has been reached on payments for crowdsourcing with no universal guidelines accepted (recent discussions on this issue can be found in Chandler & Shapiro, 2016; Goodman & Paolacci, 2017). An appropriate wage is difficult to determine given the need to balance fair wages with avoiding ethically-challenging issues concerning undue influence or practically-challenging issues with disingenuous responding due to high compensation relative to community standards. This study was designed to pay at an approximate rate of \$10/hour based on expected time of completion in pilot testing of the survey (i.e., 30 minutes). On average, our estimated time of completion closely matched the typical completion time (i.e., 31 minutes). However, wide variation was observed with many participants writing much more and for much longer than anticipated resulting in a lower-than-expected average hourly compensation (e.g., ~\$6.67/hour assuming 45 minutes or \$5/hour assuming 60 minutes). This highlights one of the challenges in designing payment structures for online studies in which completion times are hard to estimate and may vary widely person-to-person, especially for a qualitative study such as this. We believe that it is perhaps most important to consider in all of this the need to remain transparent with participants and ensuring that all expectations (e.g., expected time of completion; expected effort for a task) and incentives (e.g., payment, time to payment) are clearly articulated.

Limitations

Limitations of this study may help inform future work as well as delineate how crowdsourcing is positioned to extend and complement, but not replace, human laboratory and clinical work as well as existing work using other digital technologies. We relied on a textual analysis system that used proportional word counts generally insensitive to context. We also relied on a multilinguistic algorithm that evaluated authenticity based on empirical work of deceptive responding (Newman et al., 2003). It should be emphasized that with this information we cannot say with certainty that a response was or was not authentic. Instead, these values were used to help summarize a predictive likelihood of the extent that narratives were authentic, honest, and personal based on the pattern of linguistic style. We also relied on comparisons to more general expressive writing and other written entries rather than

those specifically related to opioid or other substance use. Although these preliminary results were promising, future research would benefit from comparing response styles and these authenticity values from individuals randomized to conditions of writing truthful and false narratives in similar online, crowdsourced settings.

Crowdsourcing also relies on online resources to collect data and therefore may fail to include individuals without Internet access. This exclusion would prove problematic for individuals with intermittent access to the Internet or difficulty in using e-banking reimbursement methods. Although some digital divides still exist, other work has also demonstrated increasing Internet and smartphone access, including among persons with current or past illicit substance use (e.g., McClure, Acquavita, Harding, & Stitzer, 2013; Strickland, Wagner, Stoops, & Rush, 2015; Tofighi, Abrantes, & Stein, 2018). The online setting also does not allow for direct contact with participants removing a possibility for clarification or direct follow-up questions. The use of structured responding cues based on textual responses are possible, albeit difficult, to code in online survey platforms. These procedures also require extensive forethought concerning likely responses and needed follow-ups. This limitation highlights the importance of rigorous study design, prepared analysis plans, and pilot testing of questions to verify whether proposed items are understood and answered as intended (e.g., through pilot focus groups).

Also relevant to note are potential limitations with generalizability and bias when using crowdsourced resources. Prior research has found that comparisons of mTurk to other forms of convenience sampling like college or community sampling suggests that mTurk participants are no more or less likely to engage in dishonest or disingenuous behavior (e.g., responding in socially acceptable ways or without paying attention) (Necka, Cacioppo, Norman, & Cacioppo, 2016). Similar research has reported that participants on mTurk, particularly those that are younger, are as or more representative of nationally representative data compared to other convenience sources like college student samples or those drawn from college town (Berinsky, Huber, & Lenz, 2012; Huff & Tingley, 2015). Nevertheless, deviations do occur from national representation in all cases of crowdsourcing with sampling that is likely biased towards those individuals who are more technological adept and/or younger. These concerns relevant to accessibility, dynamic adaptability, and generalizability highlight the importance of using crowdsourcing not as an exclusive source of information, but as an alternative, complementary source to other digital technology, laboratory, clinical, and epidemiological sources in a way that balances these tools strengths and limitations.

Conclusion

This work expands upon previous studies on mTurk that collected qualitative data (Audley, Grenier, Martin, & Ramos, 2018; McConnachie et al., 2019; Rothman, Paruk, Espensen, Temple, & Adams, 2017; Schnur et al., 2017) as well as initial efforts at formally establishing the utility of these data collection approaches (Gryzman, 2015). The current study advances this work by establishing not only the feasibility, but also providing preliminary support for the validity of qualitative data collected using crowdsourcing sources when working with people who use drugs. The flexibility of online resources

combined with access to traditionally under-represented or difficult-to-recruit participant populations provides an exciting avenue for researchers interested in efficient means to conduct mixed method and qualitative approaches to complement existing digital technologies, human laboratory, and clinic work in addiction science.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements and Funding Information

The authors declare no relevant conflicts of interest. The authors gratefully acknowledge support from the National Institute on Drug Abuse (NIDA) of the National Institutes of Health (T32 DA07209) and a Pilot Grant Award from the University of Kentucky College of Social Work. These funding agencies had no role in preparation or submission of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIDA or the University of Kentucky.

References

- Amlung M, Reed DD, Morris V, Aston ER, Metrik J, & MacKillop J (2019). Price elasticity of illegal versus legal cannabis: A behavioral economic substitutability analysis. *Addiction*, 114(1), 112–118. doi:10.1111/add.14437 [PubMed: 30194789]
- Andersen JP, Zou C, & Blosnich J (2015). Multiple early victimization experiences as a pathway to explain physical health disparities among sexual minority and heterosexual individuals. *Social Science Medicine*, 133, 111–119. [PubMed: 25864147]
- Arch JJ, & Carr AL (2017). Using mechanical turk for research on cancer survivors. *Psycho-oncology*, 26(10), 1593–1603. [PubMed: 27283906]
- Areán PA, & Gallagher-Thompson D (1996). Issues and recommendations for the recruitment and retention of older ethnic minority adults into clinical research. *Journal of Consulting Clinical Psychology*, 64(5), 875–880. [PubMed: 8916614]
- Audley S, Grenier K, Martin JL, & Ramos J (2018). Why me? An exploratory qualitative study of drinking gamers' reasons for selecting other players to drink. *Emerging Adulthood*, 6(2), 79–90.
- Barratt MJ (2012). The efficacy of interviewing young drug users through online chat. *Drug and Alcohol Review*, 31(4), 566–572. [PubMed: 22168668]
- Barratt MJ, & Maddox A (2016). Active engagement with stigmatised communities through digital ethnography. *Qualitative Research*, 16(6), 701–719.
- Bergeria CL, Huhn AS, & Dunn KE (2019). Randomized comparison of two web-based interventions on immediate and 30-day opioid overdose knowledge in three unique risk groups. *Preventive Medicine*.
- Berinsky AJ, Huber GA, & Lenz GS (2012). Evaluating online labor markets for experimental research: Amazon.Com's mechanical turk. *Political Analysis*, 20(3), 351–368. doi:10.1093/pan/mpr057
- Biernacki P, & Waldorf D (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods Research*, 10(2), 141–163.
- Blanton S, Morris DM, Prettyman MG, McCulloch K, Redmond S, Light KE, & Wolf SL (2006). Lessons learned in participant recruitment and retention: The excite trial. *Physical Therapy*, 86(11), 1520–1533. [PubMed: 17079752]
- Borodovsky JT, Marsch LA, & Budney AJ (2018). Studying cannabis use behaviors with facebook and web surveys: Methods and insights. *JMIR Public Health And Surveillance*, 4(2), e48. doi:10.2196/publichealth.9408 [PubMed: 29720366]
- Casler K, Bickel L, & Hackett E (2013). Separate but equal? A comparison of participants and data gathered via amazon's mturk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29(6), 2156–2160.

- Catalpa JM, McGuire JK, Fish JN, Nic Rider G, Bradford N, & Berg D (2019). Predictive validity of the genderqueer identity scale (gqi): Differences between genderqueer, transgender and cisgender sexual minority individuals. *International Journal of Transgenderism*, 20(2–3), 305–314.
- Chandler J, & Paolacci G (2017). Lie for a dime: When most prescreening responses are honest but most study participants are impostors. *Social Psychological Personality Science*, 8(5), 500–508.
- Chandler J, & Shapiro D (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology*, 12, 53–81. doi:10.1146/annurev-clinpsy-021815-093623
- Coderre F, Mathieu A, & St-Laurent N (2004). Comparison of the quality of qualitative data obtained through telephone, postal and email surveys. *International Journal of Market Research*, 46(3), 349–357.
- Coomber R (2011). Using the internet for qualitative research on drug users and drug markets: The pros, the cons and the progress In Fountain J, Frank VA & Korf DJ (Eds.), *Markets, methods and messages. Dynamics in European drug research* (pp. 85–103). Lengerich: Pabst Science Publishers.
- Corbin JM, & Strauss A (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 13(1), 3–21.
- Costello KL, Martin JD III, & Edwards Brinegar A (2017). Online disclosure of illicit information: Information behaviors in two drug forums. *Journal of the Association for Information Science and Technology*, 68(10), 2439–2448.
- Crump MJ, McDonnell JV, & Gureckis TM (2013). Evaluating amazon’s mechanical turk as a tool for experimental behavioral research. *PloS One*, 8(3), e57410. [PubMed: 23516406]
- Cunningham JA, Godinho A, & Bertholet N. J. B. m. r. m. (2019). Outcomes of two randomized controlled trials, employing participants recruited through mechanical turk, of internet interventions targeting unhealthy alcohol use. *BMC medical research methodology*, 19(1), e124.
- Cunningham JA, Godinho A, & Kushnir V (2017). Using mechanical turk to recruit participants for internet intervention research: Experience from recruitment for four trials targeting hazardous alcohol consumption. *BMC medical research methodology*, 17(1), e156.
- D’Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Lee KE, & Cavazos-Rehg PA (2017). Social networking online to recover from opioid use disorder: A study of community interactions. *Drug and Alcohol Dependence*, 181, 5–10. [PubMed: 29024875]
- Dugdale S, Elison S, Davies G, Ward J, & Jones M (2016). The use of digital technology in substance misuse recovery. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10(4), ePub.
- Dunn KE, Barrett FS, Herrmann ES, Plebani JG, Sigmon SC, & Johnson MW (2016). Behavioral risk assessment for infectious diseases (braid): Self-report instrument to assess injection and noninjection risk behaviors in substance users. *Drug and Alcohol Dependence*, 168, 69–75. [PubMed: 27615403]
- Dunn KE, Barrett FS, Yopez-Laubach C, Meyer AC, Hruska BJ, Sigmon SC, ... Bigelow GE (2016). Brief opioid overdose knowledge (book): A questionnaire to assess overdose knowledge in individuals who use illicit or prescribed opioids. *Journal of Addiction Medicine*, 10(5), 314. [PubMed: 27504923]
- Estellés-Arolas E, & González-Ladrón-De-Guevara F (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189–200.
- Faugier J, & Sargeant M (1997). Sampling hard to reach populations. *Journal of Advanced Nursing*, 26(4), 790–797. [PubMed: 9354993]
- Fischbacher U, & Föllmi-Heusi F (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525–547.
- Follmer DJ, Sperling RA, & Suen HK (2017). The role of mturk in education research: Advantages, issues, and future directions. *Educational Researcher*, 46(6), 329–334. doi: 10.3102/0013189x17725519
- Fusch P, & Ness LR (2015). Are we where yet? Data saturation in qualitative research. *The Qualitative Report*, 20(9), 1408–1416.

- Garcia-Romeu A, Davis AK, Erowid F, Erowid E, Griffiths RR, & Johnson MW (2019). Cessation and reduction in alcohol consumption and misuse after psychedelic use. *Journal of Psychopharmacology*, ePub ahead of print. doi:10.1177/0269881119845793
- Garfield JB, Cotton SM, Allen NB, Cheatham A, Kras M, Yücel M, & Lubman DI (2017). Evidence that anhedonia is a symptom of opioid dependence associated with recent use. *Drug and Alcohol Dependence*, 177, 29–38. [PubMed: 28551591]
- Gearhardt AN, Corbin WR, & Brownell KD (2016). Development of the yale food addiction scale version 2.0. *Psychology of Addictive Behaviors*, 30(1), 113–121. [PubMed: 26866783]
- Goodman JK, & Paolacci G (2017). Crowdsourcing consumer research. *Journal of Consumer Research*, 44(1), 196–210.
- Gryzman A (2015). Collecting narrative data on amazon's mechanical turk. *Applied Cognitive Psychology*, 29(4), 573–583.
- Harrison L, & Hughes A (1997). The validity of self-reported drug use: Improving the accuracy of survey estimates. *NIDA Research Monograph*, 167, 1–16.
- Hartsell EN, & Neupert SD (2017). Chronic and daily stressors along with negative affect interact to predict daily tiredness. *Journal of Applied Gerontology*, ePub ahead of print. 0733464817741684.
- Huff C, & Tingley D (2015). “Who are these people?” evaluating the demographic characteristics and political preferences of mturk survey respondents. *Research Politics*, 2(3), doi: 10.1177/2053168015604648.
- Huhn AS, Garcia-Romeu AP, & Dunn KE (2018). Opioid overdose education for individuals prescribed opioids for pain management: Randomized comparison of two computer-based interventions. *Frontiers in psychiatry*, 9, e34.
- Huhn AS, Meyer R, Harris J, Ayaz H, Deneke E, Stankoski D, & Bunce S (2016). Evidence of anhedonia and differential reward processing in prefrontal cortex among post-withdrawal patients with prescription opiate dependence. *Brain Research Bulletin*, 123, 102–109. [PubMed: 26711857]
- Hydock C (2018). Assessing and overcoming participant dishonesty in online data collection. *Behavior Research Methods*, 50(4), 1563–1567. [PubMed: 29274007]
- Jarmolowicz DP, Bickel WK, Carter AE, Franck CT, & Mueller ET (2012). Using crowdsourcing to examine relations between delay and probability discounting. *Behavioural Processes*, 91(3), 308–312. doi:10.1016/j.beproc.2012.09.001 [PubMed: 22982370]
- Johnson PS, Herrmann ES, & Johnson MW (2015). Opportunity costs of reward delays and the discounting of hypothetical money and cigarettes. *Journal of the Experimental Analysis of Behavior*, 103(1), 87–107. [PubMed: 25388973]
- Johnson PS, & Johnson MW (2014). Investigation of “bath salts” use patterns within an online sample of users in the united states. *Journal of Psychoactive Drugs*, 46(5), 369–378. doi: 10.1080/02791072.2014.962717 [PubMed: 25364987]
- Kaplan BA, Reed DD, Murphy JG, Henley AJ, Reed FDD, Roma PG, & Hursh SR (2017). Time constraints in the alcohol purchase task. *Experimental and Clinical Psychopharmacology*, 25(3), 186–197. doi:10.1037/pha0000110 [PubMed: 28240924]
- Kim HS, & Hodgins DC (2017). Reliability and validity of data obtained from alcohol, cannabis, and gambling populations on amazon's mechanical turk. *Psychology of Addictive Behaviors*, 31(1), 85–94. doi:10.1037/adb0000219 [PubMed: 27893213]
- Lamy FR, Daniulaityte R, Zatreh M, Nahhas RW, Sheth A, Martins SS, ... & Carlson RG (2018). “You got to love rosin: Solventless dabs, pure, clean, natural medicine.” Exploring Twitter data on emerging trends in Rosin Tech marijuana concentrates. *Drug and Alcohol Dependence*, 183, 248–252. [PubMed: 29306816]
- Lanaj K, Johnson RE, & Barnes CM (2014). Beginning the workday yet already depleted? Consequences of late-night smartphone use and sleep. *Organizational Behavior Human Decision Processes*, 124(1), 11–23.
- Landers RN, & Behrend TS (2015). An inconvenient truth: Arbitrary distinctions between organizational, mechanical turk, and other convenience samples. *Industrial Organizational Psychology*, 8(2), 142–164.

- Lane I, Waibel A, Eck M, & Rottmann K (2010). Tools for collecting speech corpora via mechanical-turk. Paper presented at the Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk.
- Lee DC, Crosier BS, Borodovsky JT, Sargent JD, & Budney AJ (2016). Online survey characterizing vaporizer use among cannabis users. *Drug and Alcohol Dependence*, 159, 227–233. [PubMed: 26774946]
- Lee YJ, Arida JA, & Donovan HS (2017). The application of crowdsourcing approaches to cancer research: A systematic review. *Cancer Medicine*, 6(11), 2595–2605. doi:10.1002/cam4.1165 [PubMed: 28960834]
- Liehr P, Marcus MT, Carroll D, Granmayeh LK, Cron SG, & Pennebaker JW (2010). Linguistic analysis to assess the effect of a mindfulness intervention on self-change for adults in substance use recovery. *Substance Abuse*, 31(2), 79–85. [PubMed: 20408059]
- Mars SG, Bourgeois P, Karandinos G, Montero F, & Ciccarone D (2014). “Every ‘never’ i ever said came true”: Transitions from opioid pills to heroin injecting. *International Journal of Drug Policy*, 25(2), 257–266. doi:10.1016/j.drugpo.2013.10.004 [PubMed: 24238956]
- Mazar N, Amir O, & Ariely D (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644.
- McClure EA, Acquavita SP, Harding E, & Stitzer ML (2013). Utilization of communication technology by patients enrolled in substance abuse treatment. *Drug and Alcohol Dependence*, 129(1–2), 145–150. doi:10.1016/j.drugalcdep.2012.10.003 [PubMed: 23107600]
- McConnachie E, Hötzel MJ, Robbins JA, Shriver A, Weary DM, & von Keyserlingk MA (2019). Public attitudes towards genetically modified polled cattle. *PloS One*, 14(5), e0216542. [PubMed: 31075123]
- Miller JD, Crowe M, Weiss B, Maples-Keller JL, & Lynam DR (2017). Using online, crowdsourcing platforms for data collection in personality disorder research: The example of amazon's mechanical turk. *Personality Disorders-Theory Research and Treatment*, 8(1), 26–34. doi:10.1037/per0000191
- Morean ME, Lipshie N, Josephson M, & Foster D (2017). Predictors of adult e-cigarette users vaporizing cannabis using e-cigarettes and vape-pens. *Substance Use Misuse*, 52(8), 974–981. [PubMed: 28323498]
- Morris V, Amlung M, Kaplan BA, Reed DD, Petker T, & MacKillop J (2017). Using crowdsourcing to examine behavioral economic measures of alcohol value and proportionate alcohol reinforcement. *Experimental and Clinical Psychopharmacology*, 25(4), 314–321. [PubMed: 28627926]
- Necka EA, Cacioppo S, Norman GJ, & Cacioppo JT (2016). Measuring the prevalence of problematic respondent behaviors among mturk, campus, and community participants. *Plos One*, 11(6), e0157732. doi:10.1371/journal.pone.0157732 [PubMed: 27351378]
- Newman ML, Pennebaker JW, Berry DS, & Richards JM (2003). Lying words: Predicting deception from linguistic styles. *Personality Social Psychology Bulletin*, 29(5), 665–675. [PubMed: 15272998]
- Parent J, Forehand R, Pomerantz H, Peisch V, & Seehuus M (2017). Father participation in child psychopathology research. *Journal of Abnormal Child Psychology*, 45(7), 1259–1270. [PubMed: 28058518]
- Paul MJ, Chisolm MS, Johnson MW, Vandrey RG, & Dredze M (2016). Assessing the validity of online drug forums as a source for estimating demographic and temporal trends in drug use. *Journal of Addiction Medicine*, 10(5), 324–330. doi:10.1097/adm.0000000000000238 [PubMed: 27466069]
- Peer E, Vosgerau J, & Acquisti A. J. B. r. m. (2014). Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior Research Methods*, 46(4), 1023–1031. [PubMed: 24356996]
- Pennebaker JW, Boyd RL, Jordan K, & Blackburn K (2015). The development and psychometric properties of LIWC 2015. Retrieved from <http://hdl.handle.net/2152/31333>
- Pruckner GJ, & Sausgruber R (2013). Honesty on the streets: A field study on newspaper purchasing. *Journal of the European Economic Association*, 11(3), 661–679.

- Rainey JC, Furman CR, & Gearhardt AN (2018). Food addiction among sexual minorities. *Appetite*, 120, 16–22. [PubMed: 28830721]
- Ramo DE, & Prochaska JJ (2012). Broad reach and targeted recruitment using facebook for an online survey of young adult substance use. *Journal of Medical Internet Research*, 14(1), e28. [PubMed: 22360969]
- Ramo DE, Thrul J, Delucchi KL, Hall S, Ling PM, Belohlavek A, & Prochaska JJ (2018). A randomized controlled evaluation of the tobacco status project, a facebook intervention for young adults. *Addiction*, 113(9), 1683–1695.
- Rhodes T, Stimson GV, Moore D, & Bourgois P (2010). Qualitative social research in addictions publishing: Creating an enabling journal environment. *The International Journal on Drug Policy*, 21(6), 441–444. [PubMed: 21051214]
- Rothman EF, Paruk J, Espensen A, Temple JR, & Adams K (2017). A qualitative study of what us parents say and do when their young children see pornography. *Academic Pediatrics*, 17(8), 844–849. [PubMed: 28450081]
- Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, ... Jinks C (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality Quantity*, 52(4), 1893–1907. [PubMed: 29937585]
- Schnur JB, Dillon MJ, Goldsmith RE, & Montgomery GH (2017). Cancer treatment experiences among survivors of childhood sexual abuse: A qualitative investigation of triggers and reactions to cumulative trauma. *Palliative Supportive Care*, 1–10.
- Sharpe Wessling K, Huber J, & Netzer O (2017). Mturk character misrepresentation: Assessment and solutions. *Journal of Consumer Research*, 44(1), 211–230.
- Sidani JE, Colditz JB, Barrett EL, Shensa A, Chu KH, James AE, & Primack BA (2019). I wake up and hit the JUUL: Analyzing Twitter for JUUL Nicotine Effects and Dependence. *Drug and Alcohol Dependence*, ePub ahead of print. doi: 10.1016/j.drugalcdep.2019.06.005
- Smith NA, Sabat IE, Martinez LR, Weaver K, & Xu S (2015). A convenient solution: Using mturk to sample from hard-to-reach populations. *Industrial Organizational Psychology*, 8(2), 220–228.
- Stewart N, Chandler J, & Paolacci G (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, 21(10), 736–748. [PubMed: 28803699]
- Stoet G (2017). Psytoolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31.
- Strickland JC, Bolin BL, Lile JA, Rush CR, & Stoops WW (2016). Differential sensitivity to learning from positive and negative outcomes in cocaine users. *Drug and Alcohol Dependence*, 166, 61–68. doi:10.1016/j.drugalcdep.2016.06.022 [PubMed: 27373186]
- Strickland JC, Hill JC, Stoops WW, & Rush CR (2019). Feasibility, acceptability, and initial efficacy of delivering alcohol use cognitive interventions via crowdsourcing. *Alcoholism: Clinical and Experimental Research*, 43(5), 888–899. doi:10.1111/acer.13987
- Strickland JC, Lile JA, & Stoops WW (2017). Unique prediction of cannabis use severity and behaviors by delay discounting and behavioral economic demand. *Behavioural Processes*, 140, 33–40. doi:10.1016/j.beproc.2017.03.017 [PubMed: 28347716]
- Strickland JC, & Stoops WW (2015). Perceptions of research risk and undue influence: Implications for ethics of research conducted with cocaine users. *Drug and Alcohol Dependence*, 156, 304–310. [PubMed: 26460141]
- Strickland JC, & Stoops WW (2018a). Evaluating autonomy, beneficence, and justice with substance-using populations: Implications for clinical research participation. *Psychology of Addictive Behaviors*, 32(5), 552–563. [PubMed: 29999330]
- Strickland JC, & Stoops WW (2018b). Feasibility, acceptability, and validity of crowdsourcing for collecting longitudinal alcohol use data. *Journal of the Experimental Analysis of Behavior*, 110(1), 136–153. doi:10.1002/jeab.445 [PubMed: 29873806]
- Strickland JC, & Stoops WW (2019). The use of crowdsourcing in addiction science research: Amazon mechanical turk. *Experimental and Clinical Psychopharmacology*, 27(1), 1–18. doi:10.1037/pha0000235 [PubMed: 30489114]

- Strickland JC, Wagner FP, Stoops WW, & Rush CR (2015). Profile of internet access in active cocaine users. *The American Journal on Addictions*, 24(7), 582–585. doi:10.1111/ajad.12271 [PubMed: 26331803]
- Sussman S, & Leventhal A (2014). Substance misuse prevention: Addressing anhedonia. *New Directions for Youth Development*, 2014(141), 45–56.
- Tausczik YR, & Pennebaker JW (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language Social Psychology*, 29(1), 24–54.
- Tofighi B, Abrantes A, & Stein MD (2018). The role of technology-based interventions for substance use disorders in primary care: A review of the literature. *The Medical clinics of North America*, 102(4), 715–731. doi:10.1016/j.mcna.2018.02.011 [PubMed: 29933825]
- Tran M, Cabral L, Patel R, & Cusack R (2017). Online recruitment and testing of infants with mechanical Turk. *Journal of Experimental Child Psychology*, 156, 168–178. doi:10.1016/j.jecp.2016.12.003 [PubMed: 28088051]
- Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, & Sonenstein FL (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science*, 280(5365), 867–873. [PubMed: 9572724]

Highlights

- We evaluate the use of crowdsourcing for collecting qualitative data
- We collected non-medical opioid to heroin/fentanyl use transition narratives
- Recruitment was feasible with sampling from diverse geographic locations
- High ratings of authenticity were observed for the provided narratives

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

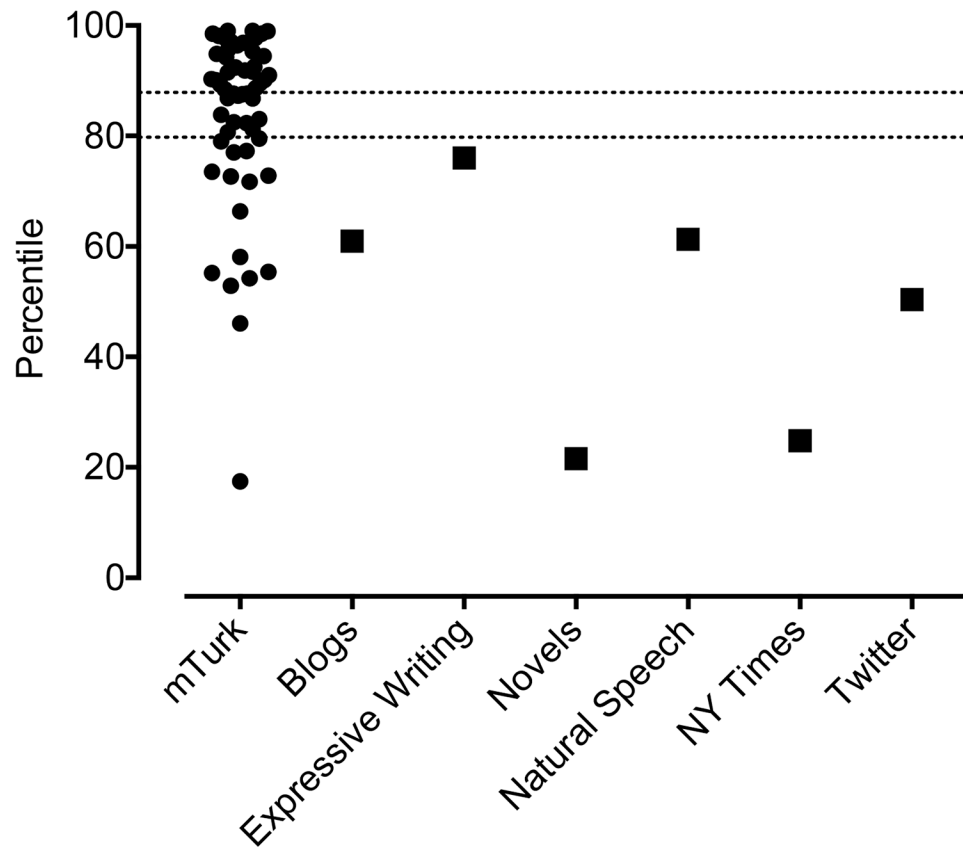


Figure 1. Authenticity Percentile for mTurk and Database Writing Sample Data.

Circles represent individual subject data for authenticity percentiles of writing narratives. Dotted lines are 95% confidence intervals surrounding the mTurk data means. Squares are point estimates for mean authenticity percentiles of writing samples from varied writing sources as documented in the LIWC 2015 handbook (Pennebaker et al. 2015).

Table 1.

Sample Demographics and Opioid Use History

	Mean (SD)/%
Demographics	
Age	34.2 (9.1)
Female	56.7%
White	80%
High School or Less	57.6%
Income (Median) ^a	\$30,001-\$40,000
Tobacco Cigarette Use	71.7%
Opioid Use History	
Ever Non-Medical Prescription Opioid Use	100.0%
Past Month Non-Medical Prescription Opioid Use	51.7%
Ever Heroin Use	95.0%
Past Month Heroin Use	45.0%
Ever Injection Drug Use	60.0%
Ever Experience Overdose	54.2%
Current MAT Treatment	33.3%
Fentanyl Use	
Yes	70.0%
Not Sure	15.0%
No	15.0%

Note. N = 60; MAT = medication-assisted treatment [methadone, buprenorphine, and naltrexone].

^aIncome was collected in \$10,000 bins and median score represents median endorsed bin.

Table 2.

Computed Text Analysis of Qualitative Responses

	Mean (95% CI)	General Database Mean	Cohen's <i>d</i> (General)	Expressive Writing Database Mean	Cohen's <i>d</i> (Expressive)
Total Word Count	377.4 (297.6, 457.1)	-	-	-	-
Authenticity (percentile)	83.8 (79.7, 87.9)	49.2	2.20***	76.0	0.50***
First-Person Singular	9.8% (9.0%, 10.7%)	5.0%	1.49***	8.7%	0.35**
Affect Language	4.2% (3.6%, 4.8%)	5.6%	0.65***	4.8%	0.27*
Positive	2.1% (1.7%, 2.5%)	3.7%	1.16***	2.6%	0.34*
Negative	2.1% (1.7%, 2.4%)	1.8%	0.18	2.1%	0.04

Note. General database mean and expressive writing database mean data are from the LIWC 2015 handbook (Pennebaker et al. 2015). Effect size estimates are Cohen's *d* from a one-sample *t*-test relative to sample mean data.

*
 $p < .05$;

**
 $p < .01$;

 $p < .001$