

UCSF

UC San Francisco Previously Published Works

Title

Comparing Amazon Mechanical Turk with unpaid internet resources in online clinical trials

Permalink

<https://escholarship.org/uc/item/3568344x>

Authors

Bunge, Eduardo

Cook, Haley M

Bond, Melissa

et al.

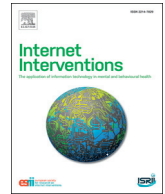
Publication Date

2018-06-01

DOI

10.1016/j.invent.2018.04.001

Peer reviewed



Comparing Amazon Mechanical Turk with unpaid internet resources in online clinical trials



Eduardo Bunge*, Haley M. Cook, Melissa Bond, Rachel E. Williamson, Monique Cano, Alinne Z. Barrera, Yan Leykin, Ricardo F. Muñoz

Palo Alto University, i4Health, United States

ARTICLE INFO

Keywords:

Internet interventions
Recruitment
Amazon Mechanical Turk
Clinical trials
Depression

ABSTRACT

Internet interventions face significant challenges in recruitment and attrition rates are typically high and problematic. Finding innovative yet scientifically valid avenues for attaining and retaining participants is therefore of considerable importance. The main goal of this study was to compare recruitment process and participants characteristics between two similar randomized control trials of mood management interventions. One of the trials (Bunge et al., 2016) was conducted with participants recruited from Amazon's Mechanical Turk (AMT), and the other trial recruited via Unpaid Internet Resources (UIR).

Methods: The AMT sample (Bunge et al., 2016) consisted of 765 adults, and the UIR sample (recruited specifically for this study) consisted of 329 adult US residents. Participants' levels of depression, anxiety, confidence, motivation, and perceived usefulness of the intervention were assessed. The AMT sample was financially compensated whereas the UIR was not.

Results: AMT yielded higher recruitment rates per month ($p < .05$). At baseline, the AMT sample reported significantly lower depression and anxiety scores ($p < .001$ and $p < .005$, respectively) and significantly higher mood, motivation, and confidence (all $p < .001$) compared to the UIR sample. AMT participants spent significantly less time on the site ($p < .05$) and were more likely to complete follow-ups than the UIR sample ($p < .05$). Both samples reported a significant increase in their level of confidence and motivation from pre- to post-intervention. AMT participants showed a significant increase in perceived usefulness of the intervention ($p < .0001$), whereas the UIR sample did not ($p = .1642$).

Conclusions: By using AMT, researchers can recruit very rapidly and obtain higher retention rates; however, these participants may not be representative of the general online population interested in clinical interventions. Considering that AMT and UIR participants differed in most baseline variables, data from clinical studies resulting from AMT samples should be interpreted with caution.

1. Introduction

Internet interventions are increasing in popularity and use within the health field (Andersson and Cuijpers, 2009a; Baumeister et al., 2014; Richards et al., 2015). They have proven to be more cost effective than therapy as usual and less stigmatizing for individuals who need mental health care; they offer convenience in a fast paced society, are “non-consumable”, and have the potential to reduce health disparities worldwide (Leykin et al., 2014; Muñoz, 2010; Ritterband and Tate, 2009). However, Internet interventions face two significant challenges: recruitment and attrition.

Recruitment is one of the most difficult components of a clinical research study (Ashery and McAuliffe, 1992; Patel, 2003). Recruiting participants and conducting studies on the Internet has considerable

advantages in comparison to laboratory-based research (Hoerger, 2010). Benefits of Internet-based research include the ability to recruit larger and more diverse samples, automation of data entry, and reduction in both staff time and financial costs (Gosling and Mason, 2015; Gosling et al., 2004; Hoerger, 2010; Nosek et al., 2002).

Online recruitment methods include: Internet advertising (e.g. Google AdWords), organic online traffic (e.g., following links displayed on other websites or in search results), and paid participant platforms (e.g. Amazon's Mechanical Turk [AMT]). Internet advertising offers an expensive, but relatively effortless option for reaching a vast number of people (Miller and Sønderlund, 2010), and can be tailored specifically to a target population (see Graham et al., 2012; Gross et al., 2014). For instance, Google AdWords campaigns present users ads based on related word searches; if the ad is clicked, the individual is taken to the

* Corresponding author at: Institute for International Internet Interventions for Health (i4Health), Palo Alto University, 1791 Arastradero Rd., Palo Alto, CA 94304, United States.
E-mail address: ebunge@paloinstitution.edu (E. Bunge).

study's website. Internet advertising can likewise be done through Social Media (e.g., Facebook), which may be effective in attaining hard to reach (Whitaker et al., 2017) and diverse samples with successful retention rates. Facebook, being one of the most widely used social networks in the world (Kayrouz et al., 2016), is increasingly being used as a research recruitment tool.

AMT is an online service that connects researchers, and developers with individuals who perform human intelligence tasks (e.g., completing surveys, doing web-based experiments, etc.) for compensation (Buhmester et al., 2011; Mason and Suri, 2012; Paolacci et al., 2010). AMT offers a rapid, inexpensive approach for recruiting large and diverse samples compared to more traditional recruitment methods (Mason and Suri, 2012; Paolacci et al., 2010). Given the small payment from each completed task, it is unlikely that AMT workers use the site as a primary source of income; it is therefore less likely that financial motivation may impact research results (Buhmester et al., 2011; Paolacci et al., 2010). Studies that have examined the quality and validity of data gathered via AMT have concluded that low compensation rates do not affect data quality (Buhmester et al., 2011) therefore making it a viable option for recruiting research participants. Several studies have examined the validity of AMT-generated data for clinical studies (Cunningham et al., 2017; Shapiro et al., 2013). Cunningham et al. (2017) observed a rapid recruitment rate, and Shapiro et al. (2013) found that data collection (on mental health disorders) was fast and data quality was high; but highlighted that precautions measures should be implemented during recruitment. However, none of these studies directly compared a sample recruited through AMT to one recruited through other Internet recruitment methods.

Attrition rates in Internet intervention research are typically high and problematic (Andersson and Cuijpers, 2009b; Eysenbach, 2005). Many individuals who take part in Internet interventions studies visit the website only once (Clarke et al., 2002; Leykin et al., 2014), and as many as 60% of participants never complete a follow-up (Andersson and Cuijpers, 2009a, 2009b; Eysenbach, 2005) Attrition affects the representativeness of the original sample and the generalizability of the study's results (Hoerger, 2010). Hoerger (2010), noted that participants tended to drop out of an Internet-based study after reading the consent form or at the start of a study, which was likely not due to fatigue or possible fear of harm. Online participants are perhaps less motivated for treatment than those who take the time to travel to sites conducting face-to-face studies (Eysenbach, 2005).

Bunge et al. (2016) proposed that brief Internet interventions (BII) could help reduce the attrition problems. Because BII can be tested more rapidly than full-scale treatment protocols, intervention improvements can be incorporated faster, leading to improved participant experiences and better outcomes, which may consequently increase the likelihood of retaining participants for both the intervention and the follow-up. Indeed, a randomized controlled trial of different brief mood management strategies using AMT participants produced one-week follow-up completion rates of 60.65% (Bunge et al., 2016). The follow-up completion rates were quite high for an unsupported (no human contact) Internet intervention study; however, considering that it was a paid study, it is not clear whether a similar retention rate could be obtained with a non-paid Internet sample. Thus, the main goal of the present study is to compare the recruitment process, baseline characteristics of the participants, time spent on the site, follow-up completion rates, and immediate post-test outcome differences between a study previously done by Bunge et al. (2016) and an identical study recruiting participants through Unpaid Internet Resources (UIR).

2. Method

2.1. Participants

The total sample consisted of 1187 participants: 765 recruited from the AMT and 329 – from UIR. The AMT sample was described in Bunge

et al. (2016). Eligibility criteria for Bunge et al. (2016) and the current study was the same, and included that participants be at least 18 years of age and provide informed consent. The Institutional Review Board (IRB) at Palo Alto University approved this study. The trial was registered at clinicaltrials.gov, ref. number: NCT02748954.

2.2. Measures

2.2.1. Demographic questionnaire

Participants indicated their gender, age, country of residence, and zip (postal) code.

2.2.2. Patient health questionnaire-9 (PHQ-9; Kroenke and Spitzer, 2002)

Is the widely used and well-validated 10-item measure of depression symptoms.

2.2.3. Generalized Anxiety Disorder questionnaire (GAD-7; Spitzer et al., 2006)

Is a commonly used 7-item measure of symptoms of generalized anxiety disorder plus an item related to global functioning.

2.2.4. Ancillary questions

Participants were asked to answer four Likert-type questions: - "How would you describe your mood in the last 2 weeks?" with responses ranging from 0 = *Extremely Negative* to 9 = *Extremely Positive*.

- "How motivated are you to do something to improve your mood?" with responses ranging from 0 = *Not at all* to 10 = *Extremely*.

- "How confident are you that you are able to do something to improve your mood?"

with responses ranging from 0 = *Not at all* to 10 = *Extremely*.

- "Before you see the ideas we will be sharing with you, how likely do you think they will be useful?" with responses ranging from 0 = *Not at all* to 10 = *Extremely*. At post-test and follow-up, instead of this question, participants were asked: "How useful was this intervention?", with responses ranging from 0 = *Not at all* to 10 = *Extremely*.

2.2.5. Attention questions

Participants in the AMT sample only were asked to complete two attention questions. The first one – Paying Attention – required the participant to respond to a single item: "Please verify your place in this survey by selecting 'very much'" with three answer choices: 1. *Not at all*, 2. *Somewhat*, 3. *Very much*. The second question – Subjective Attention – inquired about self-reported carefulness in responses, "We would like to know how carefully you completed this survey. You will be paid regardless of your answers" and given a 5-point multiple-choice selection with the answer choices: 1. *Very carefully*, 2. *Carefully*, 3. *As well as I answer most surveys*, 4. *Quickly*, 5. *As quickly as I could*.

2.3. Procedures

2.3.1. Recruitment procedures

Recruitment of AMT participants is described in Bunge et al. (2016). The UIR participants were primarily recruited through free press releases, list serves, free websites recruiting participants for research studies, and through social media. More specifically, participants were recruited from Hanover (online database of online interventions - 28.2% of participants were recruited in this manner), Craigslist (17.8%), Depression and Bipolar Support Alliance listserv (16.6%), National Latina/o Psychological Association listserv (13%), Reddit (10%), and other avenues (3%). The description of the study was similar for both samples, with the sole difference being the description of the incentives – the AMT study description stated that participants would be compensated, whereas the UIR description stated that the study was uncompensated. The compensation for the AMT participants was US\$0.10 for completing baseline, intervention and immediate post-test, and another US\$0.10 for completing the one-week follow-up.

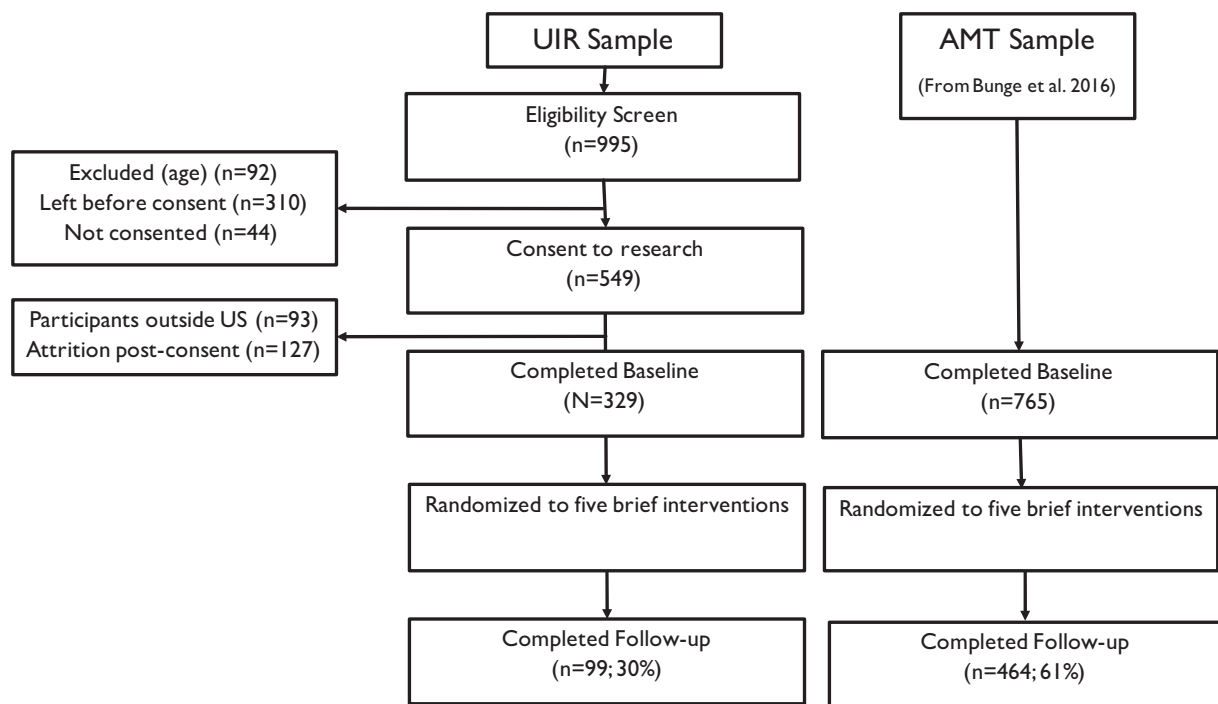


Fig. 1. Progression of participants recruited from Amazon Mechanical Turk (AMT) and Unpaid Internet Resources (UIR). The Consort flow chart represents the AMT sample from Bunge et al. (2016) and the UIR sample.

Examples of UIR study advertisements were: “Improve your mood on your own. Join Free Palo Alto Univ. study to improve your mood online + [link to the site]”; “Ready to improve your mood? Join Free Palo Alto Univ. study to improve your mood on your own + [link to the study site]”; “Healthy Mood Online Study. Researchers at Palo Alto University are conducting a free online study to examine whether very brief interventions help people improve their mood. To join the study, which takes 10-20 minutes to complete online, plus a 5-minute survey a week later, go to [link to study site]”. As described in Bunge et al. (2016), the AMT study advertisement included information such as who was the requester, expiration date, time allocated, reward, description (“improve your mood on your own”), and keywords (improve mood, psychology, feel better).

2.3.2. Study procedures

After being redirected to the study website, participants were asked to indicate their age (to establish eligibility). Eligible participants were then presented with the option of either providing consent for researchers to use their data in research reports or using the site without providing such consent; only data from those participants who consented to have their data used for research are described below. Study procedures are described elsewhere (Bunge et al., 2016). Briefly, consenting participants were asked to complete demographic questions, the four ancillary questions, the PHQ-9, and the GAD-7. Participants were then randomized to the following conditions: (1) Thoughts (increasing helpful thoughts); (2) Activities (increasing activity level); (3) Assertiveness (tips for communicating assertively); (4) Sleep hygiene (a description on how sleep can affect mood); and (5) Own Methods (participants were asked to identify four of their own personal strategies that have helped them improve their mood in the past). After completing the intervention, the AMT participants responded to the questions about their motivation, confidence, and usefulness of the intervention, as well as both “attention” questions. One week after completing the intervention, participants were sent a follow-up survey with the ancillary questions, PHQ-9, and the GAD-7. At the end of the study, all participants were provided a link that allowed them to access all five of the intervention conditions. Additionally, participants were

provided other resources, such as e-couch and the USA Crisis Call Center.

The AMT sample ($n = 765$) were recruited over a period of one month (May to June 2015). Total recruitment cost for the AMT sample was US\$122.90 plus staff time to manage payments. The UIR participants ($n = 329$) was recruited over a period of 19 months (July 2014 to February 2016) and received no compensation; total recruitment cost was therefore US\$0, plus staff time to post study advertisements on the Web.

2.4. Statistical analysis

Differences between the AMT and UIR samples vis-à-vis demographics, symptomatology, time spent on the site, and follow-up rates were examined via independent samples *t*-tests and chi-square tests, as appropriate. Mixed effects ANOVAs were used to detect differences between the two samples in regard to changes from pre- to post-intervention in ratings of motivation to change one's mood, confidence in one's ability to change one's mood, and perceived usefulness of the assigned intervention.

3. Results

3.1. Recruitment process

Of the 765 AMT participants who completed the intervention, 464 participants (61%) completed the one-week follow-up survey (see Bunge et al., 2016). Of the 995 UIR participants who entered the survey, 92 participants were immediately screened out due to their age (under 18 years old). Of the 903 UIR participants who were eligible, 44 did not consent to have their data used, but continued with the intervention; 549 consented to participate in the research study. Of those who consented, 127 participants dropped out before receiving the intervention. Of the remaining 422 participants who completed baseline, 93 reported living outside of the United States and were removed from the sample before data analysis to make the two samples more comparable. Thus, the final UIR sample consisted of 329 participants who

Table 1
Comparison of UIR (*n* = 329) and AMT (*n* = 765) samples.

	UIR	Amazon Mechanical Turk
Recruitment rate (p's/month)	~17/month	~765/month
Total cost per participant	\$0.00	\$0.20
Gender (female)	76.22%	69.15% ^a
<i>M</i> _{age}	33.92 (14.3)	35.9 (8.7) ^a
Baseline PHQ9	11.05 (7.25)	8.73 (6.8) ^a
Baseline MDE	35.87%	23.92% ^a
Baseline GAD7	8.29 (5.58)	7.11 (5.82) ^a
Baseline mood rating	4.29 (2.0)	5.15 (1.94) ^a
Baseline motivation	6.48 (2.51)	7.04 (2.10) ^a
Baseline confidence	5.90 (2.64)	6.69 (2.34) ^a
Baseline usefulness	5.78 (2.11)	6.03 (2.03)
Completion time	8.2 min	6.0 min ^a
Follow-up rate	30.09%	60.65% ^a

Note: Mean and Standard Deviation presented for non-percentage and dollar values.

MDE: Screened positive for major depressive episode on the PHQ-9; GAD: Screened positive for generalized anxiety disorder.

^a Significantly different (*p* < .05) from UIR Sample.

completed the baseline survey, of whom 99 participants (30%) completed the one-week follow-up survey (See Fig. 1).

3.2. Sample characteristics at baseline

Baseline characteristics, enrollment figures, and follow-up completion rates are presented in Table 1. To facilitate the comparison between both samples, we present numbers from the published AMT data (Bunge et al., 2016) together with the current UIR sample. The AMT sample consisted of 765 adults residing in the United States (*M*_{age} = 35.9; *SD* = 8.7); 69% female. The sample of participants recruited via UIR consisted of 329 adults residing in the United States (*M*_{age} = 33.92; *SD* = 14.30); 76% female. On average, the UIR sample was younger and consisted of a higher percentage of female participants compared to the AMT sample (*p* < .05). At baseline, the AMT sample reported significantly lower PHQ-9 scores (*t*(1092) = 5.06, *p* < .001,

d = 0.33) and GAD-7 scores (*t*(1092) = 3.08, *p* < .005, *d* = 0.20). Further, the AMT sample reported significantly higher subjective mood ratings (*t*(1092) = 6.61, *p* < .001, *d* = 0.44), motivation (*t*(534) = 3.52, *p* < .001, *d* = 0.25), and confidence (*t*(558) = 4.69, *p* < .001, *d* = 0.32); no significant difference between the samples regarding expected usefulness of the intervention was observed (*t*(1092) = 1.87, *p* = .062, *d* = 0.12).

3.3. Activity completion

Because the AMT sample was compensated, two attention questions were used to determine whether participants accurately responded to items. All participants in the AMT sample (*n* = 765) completed the Paying Attention question, and of those, 98% (*n* = 750) answered the attention-check question correctly. Regarding the Subjective Attention question, 87% (*n* = 663) reported completing the survey either “carefully” or “very carefully” compared to other AMT questionnaires they have completed in other studies. Completion time for baseline, intervention, and immediate post-test indicated that AMT participants spent, on average, 6.0 min (*SD* = 3.8 min) on the study, whereas the UIR participants spent 8.2 min (*SD* = 4.8 min) (*p* > 0.05).

3.4. Outcome differences between AMT and UIR

Due to the significant differences between samples at baseline, gender, age, anxiety, and depression status were included as covariates in the mixed effects ANOVA models. Regarding the confidence in one's ability to change their mood, the results indicated a main effect of sample (*F*(1,1086) = 29.07, *p* < .0001) and time (*F*(1,1064) = 66.38, *p* < .0001, but no significant interaction, suggesting that the AMT sample reported higher levels of confidence overall, but both samples reported a similar increase in confidence from pre-intervention to post-intervention. Similarly, regarding motivation to improve their mood, main effect of sample (*F*(1,1082) = 18.69, *p* < .0001) and time (*F*(1,1054) = 24.98, *p* < .0001) was observed, again with no significant interaction. Although AMT participants reported overall higher motivation, both samples reported an increase in motivation from pre-

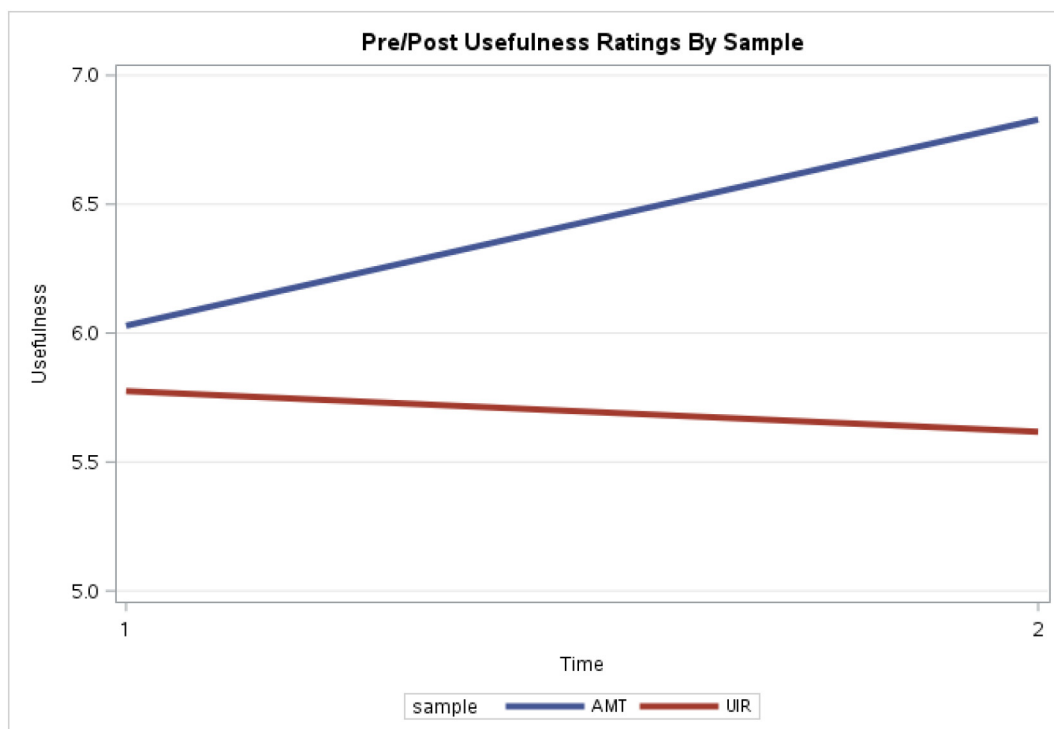


Fig. 2. Usefulness ratings by sample at pre-intervention and post-intervention. AMT: Amazon Mechanical Turk; UIR: Unpaid internet resources.

intervention to post-intervention. Regarding usefulness, the models indicated main effects of sample ($F(1,1089) = 22.80, p < .0001$) and time ($F(1,1042) = 19.66, p < .0001$), as well as an interaction of time and sample ($F(1,1042) = 43.92, p < .0001$). Thus, AMT participants demonstrated a significant increase in perceived usefulness from pre-intervention to post-intervention ($M_{\text{Diff}} = 0.806, t(1038) = 10.21, p < .0001$), but the UIR sample did not show a significant change ($M_{\text{Diff}} = -0.160, t(1043) = 1.30, p = .19$) (See Fig. 2).

4. Discussion

With the increasing popularity of Internet interventions (Muñoz, 2010a; Ritterband and Tate, 2009; Leykin et al., 2014), the need to address the challenges of recruitment and attrition is gaining in importance. Paid participant platforms, such as AMT, offer a potential solution for addressing recruitment problems and decreasing attrition rates. The current study compared recruitment processes, baseline characteristics, outcomes, completion time, and follow-up completion rates of an AMT sample versus a sample recruited through UIR.

The recruitment process yielded interesting differences. Recruitment through AMT was more efficient, as evidenced by much higher recruitment rates per month (765 participants per month for AMT vs. 17 for UIR). While the AMT participants were compensated, there was no monetary compensation for the UIR participants; the cost to recruit UIR participants (setting aside staff time, which is also needed for AMT recruitment) is effectively zero. Therefore, AMT participant recruitment allowed for a faster recruitment rate, but it also required a higher financial cost. It should be noted that, although we have been able to identify a number of Internet resources that allow us to recruit for free, given the time needed to recruit participants (19 months for 329 participants), for many researchers that would in fact be a costlier option considering the staff time and other resources needed to sustain a lengthier recruitment period. Recruitment findings from this study are consistent with existing literature reports that AMT offers a rapid and relatively inexpensive approach for recruitment (Mason and Suri, 2012).

Baseline characteristics of the two samples differed across most variables. Specifically, the UIR sample was generally younger and consisted of a higher percentage of female participants, compared to the AMT sample. Additionally, the UIR sample reported higher levels of anxiety and depression symptoms at baseline and significantly lower subjective mood (at a moderate effect size), motivation, and confidence ratings. Previous studies using ATM suggest that the collected data is comparable to more traditional subject pools (Buhrmester et al., 2011; Paolacci et al., 2010), but studies on clinical populations have raised concerns about representativeness of the ATM samples, with some research calling for greater precautions when using such samples (Shapiro et al., 2013). Interestingly, neither Shapiro et al. (2013) nor Cunningham et al. (2017) compared participants recruited through AMT to those recruited via other Internet recruitment methods. To our knowledge, the current study is the first comparing both type of samples; our results seem to agree with the previously voiced concerns about the representativeness of the AMT participants for clinical studies.

When the time that both samples of participants had spent completing baseline, interventions and immediate post-test were compared, the following differences appeared. AMT participants completed baseline, intervention, and immediate post-test significantly faster than those in the UIR sample. Such difference may be attributed to the fact that AMT participants have more practice in answering questions online. However, it is likely that AMT participants are less interested in a mood management intervention than those who actively seek for such intervention without expecting to be paid for completing it. Interestingly, 98% of AMT participants answered the attention check correctly, and most participants reported completing the survey either “carefully” or “very carefully” relative to other AMT surveys they’d

completed. Previous research suggests that the quality of AMT data may not be affected by a small monetary compensation (Buhrmester et al., 2011; Paolacci et al., 2010), and in our study the vast majority of participants passed the attention check; nonetheless, the differences in completion time suggest that AMT participants go through such studies differently than regular Internet users.

Regarding follow up completions rates, the AMT sample yielded a significantly higher follow up rate than the UIR sample. It is possible that this difference was driven by the monetary compensation: AMT workers were compensated twice (after completing baseline, intervention and immediate post-test, and after completing follow up), whereas UIR participants were not compensated at all. Although UIR samples can also be compensated in Internet studies, payment methods are usually more complicated for the participants (and researchers) than for the AMT participants, and the risk of bots represents an additional complication (Prince et al., 2012). Additionally, it is possible that AMT participants spend more time with computers than the UIR sample, which may facilitate the follow-up completion process. Overall, considering that Internet studies struggle with attrition, this represents an advantage of AMT.

The finding that AMT participants rated the intervention as more useful than UIR participants is counterintuitive considering the hypothesis that UIR participants were actively seeking out a mood improvement tool and could thus perceive it as more useful. Since this result did not come out as expected, additional concerns are raised regarding the way that AMT participants perceive interventions within clinical research studies.

4.1. Limitations and future directions

While this study was intended to compare AMT and UIR, the design does not allow to determine whether the differences between the samples were due to the compensation or to the familiarity that AMT participants may have with Internet-based tasks. Additionally, these findings may not generalize to samples that could be recruited through paid online sites, which have shown to be useful in recruiting hard to reach (Whitaker et al., 2017), diverse samples, and with successful retention rates (Fenner et al., 2012; Ramo and Prochaska, 2012; Ryan, 2013). Future studies analyzing the effect of compensation on outcomes may benefit from a 2×2 design, wherein participants recruited from UIR and AMT are randomized to either a higher or lower compensation group, to tease apart the effects of compensation and familiarity.

Because several baseline characteristics were significantly different, future studies aiming to compare participant characteristics across platforms may benefit from advertising the study with more narrow and precise inclusion criteria. Additionally, as the intervention was intended to be brief, several demographic questions were simplified or not asked, such as education level. It is possible that education level may explain some of differences between samples, especially the response completion time. Finally, the larger study tested brief automated Internet interventions for mood management and the findings may not generalize to more complex interventions, such as multi-lesson online treatments and/or to interventions delivered with human support.

5. Conclusions

Internet interventions, generally proceed through three stages: development, evaluation, and dissemination. Consistent with Cunningham et al. (2017), we conclude that AMT may represent a viable platform for initial development and pilot testing of interventions, as well as for usability and feasibility studies, as the service allows for fast, inexpensive recruitment relative to many UIR options. However, UIR recruitment options are more likely to reach and/or include individuals that represent the intended audience of Internet interventions – individuals who are internally motivated to seek out and use such interventions. Additionally, AMT participants may report the

intervention as more useful than UIR participants. Therefore, at the evaluation phase of the intervention, complementary avenues, such as online and social media sites, should be utilized in order to obtain more representative samples and, consequently, more representative findings. Finally, if the goal of Internet Intervention is to reach as many individuals as possible, the dissemination phase must necessarily go beyond paid platforms.

While many researchers attempt to address all three phases of an intervention at once, online interventions may have more success reaching their population of interest, increasing participant retention, and improving outcome data by utilizing a stepwise model and multiple avenues of recruitment instead. Paid platforms such as AMT may offer a preliminary testing ground for iterative improvements and enhancements of Internet interventions, making them more pleasurable and rewarding for users, and consequently increasing the chances of helping the intended audiences. However, data from clinical studies with AMT samples should be interpreted with caution.

Conflict of interest

The authors report no conflict of interest.

References

- Andersson, G., Cuijpers, P., 2009a. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cogn. Behav. Ther.* 38 (4), 196–205.
- Andersson, G., Cuijpers, P., 2009b. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cogn. Behav. Ther.* 38 (4), 196–205. <http://dx.doi.org/10.1080/16506070903318960>.
- Ashery, R.S., McAuliffe, W.E., 1992. Implementation issues and techniques in randomized trials of outpatient psychosocial treatments for drug abusers: recruitment of subjects. *Am. J. Drug Alcohol Abuse* 18 (3), 305–329.
- Baumeister, H., Reichler, L., Munzinger, M., Lin, J., 2014. The impact of guidance on internet-based mental health interventions — a systematic review. *Internet Intervent.* 1 (4), 205–215. <http://dx.doi.org/10.1016/j.invent.2014.08.003>.
- Buhrmester, M., Kwang, T., Gosling, S.D., 2011. Amazon's mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspect. Psychol. Sci.* 6 (1), 3–5. <http://dx.doi.org/10.1177/1745691610393980>.
- Bunge, E.L., Williamson, R.E., Cano, M., Leykin, Y., Muñoz, R.F., 2016. Mood management effects of brief unsupported internet interventions. *Internet Interventions* 5, 36–43.
- Clarke, G., Reid, D.E., O'Connor, E., DeBar, L.L., Kelleher, C., Lynch, F., Nunley, S., 2002. Overcoming depression on the internet (ODIN): a randomized controlled trial of an internet depression skills intervention program. *J. Med. Internet Res.* 4 (3) Retrieved from. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1761939/>.
- Cunningham, J.A., Godinho, A., Kushnir, V., 2017. Can Amazon's Mechanical Turk be used to recruit participants for internet intervention trials? A pilot study involving a randomized controlled trial of a brief online intervention for hazardous alcohol use. *Internet Intervent.* 10 (Supplement C), 12–16. <http://dx.doi.org/10.1016/j.invent.2017.08.005>.
- Eysenbach, G., 2005. The law of attrition. *J. Med. Internet Res.* 7 (1), e11. <http://dx.doi.org/10.2196/jmir.7.1.e11>.
- Fenner, Y., Garland, S.M., Moore, E.E., Jayasinghe, Y., Fletcher, A., Tabrizi, S.N., ... Wark, J.D., 2012. Web-based recruiting for Health Research using a social networking site: an exploratory study. *J. Med. Internet Res.* 14 (1), e20. <http://dx.doi.org/10.2196/jmir.1978>.
- Gosling, S.D., Mason, W., 2015. Internet research in psychology. *Annu. Rev. Psychol.* 66 (1), 877–902. <http://dx.doi.org/10.1146/annurev-psych-010814-015321>.
- Gosling, S.D., Vazire, S., Srivastava, S., John, O.P., 2004. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *Am. Psychol.* 59 (2), 93–104. <http://dx.doi.org/10.1037/0003-066X.59.2.93>.
- Graham, A.L., Fang, Y., Moreno, J.L., Streiff, S.L., Villegas, J., Muñoz, R.F., ... Vallone, D.M., 2012. Online advertising to reach and recruit Latino smokers to an internet cessation program: impact and costs. *J. Med. Internet Res.* 14 (4). <http://dx.doi.org/10.2196/jmir.2162>.
- Gross, M.S., Liu, N.H., Contreras, O., Muñoz, R.F., Leykin, Y., 2014. Using Google AdWords for international multilingual recruitment to Health Research websites. *J. Med. Internet Res.* 16 (1), e18. <http://dx.doi.org/10.2196/jmir.2986>.
- Hoerger, M., 2010. Participant dropout as a function of survey length in internet-mediated university studies: implications for study design and voluntary participation in psychological research. *Cyberpsychol. Behav. Soc. Netw.* 13 (6), 697–700.
- Kayrouz, R., Dear, B.F., Karin, E., Titov, N., 2016. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. *Internet Interventions* 4, 1–10.
- Kroenke, K., Spitzer, R.L., 2002. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatr. Ann.* 32, 509–515.
- Leykin, Y., Muñoz, R.F., Contreras, O., Latham, M.D., 2014. Results from a trial of an unsupported internet intervention for depressive symptoms. *Internet Intervent.* 1 (4), 175–181. <http://dx.doi.org/10.1016/j.invent.2014.09.002>.
- Mason, W., Suri, S., 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behav. Res. Methods* 44 (1), 1–23. <http://dx.doi.org/10.3758/s13428-011-0124-6>.
- Miller, P.G., Sønderlund, A.L., 2010. Using the internet to research hidden populations of illicit drug users: a review. *Addiction* 105 (9), 1557–1567.
- Muñoz, R.F., 2010. Using evidence-based internet interventions to reduce health disparities worldwide. *J. Med. Internet Res.* 12 (5), e60. <http://dx.doi.org/10.2196/jmir.1463>.
- Nosek, B.A., Banaji, M.R., Greenwald, A.G., 2002. E-research: ethics, security, design, and control in psychological research on the internet. *J. Soc. Issues* 58 (1), 161–176.
- Patel, M.X., 2003. Challenges in recruitment of research participants. *Adv. Psychiatr. Treat.* 9 (3), 229–238. <http://dx.doi.org/10.1192/apt.9.3.229>.
- Paolacci, G., Chandler, J., Ipeirotis, P.G., 2010. Running experiments on Amazon Mechanical Turk. *Judgm. Decis. Mak.* 5 (5), 411–419.
- Prince, K.R., Litovsky, A.R., Friedman-Wheeler, D.G., 2012. Research forum-internet-mediated research: beware of bots. *Behav. Therapist* 35 (5), 85.
- Ramo, D.E., Prochaska, J.J., 2012. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. *J. Med. Internet Res.* 14 (1), e28. <http://dx.doi.org/10.2196/jmir.1878>.
- Richards, D., Richardson, T., Timulak, L., McElvaney, J., 2015. The efficacy of internet-delivered treatment for generalized anxiety disorder: a systematic review and meta-analysis. *Internet Intervent.* 2 (3), 272–282. <http://dx.doi.org/10.1016/j.invent.2015.07.003>.
- Ritterband, L.M., Tate, D.F., 2009. The science of internet interventions. *Ann. Behav. Med.* 38 (1), 1–3. <http://dx.doi.org/10.1007/s12160-009-9132-5>.
- Ryan, G.S., 2013. Online social networks for patient involvement and recruitment in clinical research. *Nurs. Res.* 21 (1), 35–39.
- Shapiro, D.N., Chandler, J., Mueller, P.A., 2013. Using Mechanical Turk to study clinical populations. *Clin. Psychol. Sci.* 1 (2), 213–220. <http://dx.doi.org/10.1177/2167702612469015>.
- Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch. Intern. Med.* 166, 1092–1097.
- Whitaker, C., Stevelink, S., Fear, N., 2017. The use of Facebook in recruiting participants for Health Research purposes: a systematic review. *J. Med. Internet Res.* 19 (8), e290. <http://dx.doi.org/10.2196/jmir.7071>.