How one small text change in a study document can impact recruitment rates and follow-up completions

Alexandra Godinho[a], Christina Schell[a], John A. Cunningham[a,b,c,*]

[a] Institute of Mental Health and Policy Research, Centre for Addiction and Mental Health, Toronto, Canada
[b] Department of Psychiatry, University of Toronto, Toronto, Canada
[c] Australian National University, Canberra, Australia

A B S T R A C T

Background: The validity and reliability of longitudinal research is highly dependent on the recruitment and retention of representative samples. Various strategies have been developed and tested for improving recruitment and follow-up rates into health-behavioural research, but few have examined the role of linguistic choices and study document readability on participation rates. This study examined the impact of one small text change, assigning an inappropriate or grade-8 reading level password for intervention access, on participation rates and attrition in an online alcohol intervention trial.
Methods: Participants were recruited into an online alcohol intervention study using Amazon's Mechanical Turk via a multi-step recruitment process which required participants to log into a study portal using a pre-assigned password. Passwords were qualitatively coded as grade-8 and/or inappropriate for use within a professional setting. Separate logistic regressions examined which demographic, clinical characteristics, and password categorizations were most strongly associated with recruitment rates and follow-up completions.
Results: Inappropriate passwords were a barrier for recruitment among participants with post-secondary education as compared to those with less education ($p = 0.044$), while grade-8 passwords appeared to significantly facilitate the completion of 6-month follow-ups ($p = 0.005$).
Conclusions: Altogether, these findings suggest that some linguistic choices may play an important role in recruitment, while others, such as readability, may have longer-term effects on follow-up rates and attrition. Possible explanations for the findings, as well as, sample selection biases during recruitment and follow-up are discussed. Limitations of the study are stated and recommendations for researchers are provided.
Trial registration: ClinicalTrials.gov NCT02977026. Registered 27 Nov 2016.

1. Introduction

The validity and reliability of evidence generated from longitudinal designs is highly dependent on the recruitment *and* retention of sufficiently large and representative participant samples. Attrition at any point of a repeated-measures design (i.e. at any stage of recruitment or follow-up) is known to compromise study power; but more importantly, drop-outs that do not occur at random have been found to result in sample selection biases, and unreliable and unreproducible results (Capaldi and Patterson, 1987; Kristman et al., 2004; Barry, 2005; von Allmen et al., 2015). While various strategies and statistical methods have been developed for addressing under-powered studies (e.g. pooled analyses) and missing data (e.g. multiple imputations), most researchers would agree that collecting sufficient data points for the targeted sample at each study phase is preferable. Nonetheless,

recruiting participants for studies involving hard-to-reach samples, such as community samples of substance users, continues to be a hurdle for many research teams.

Several researchers have explored and evaluated different methods for improving recruitment into longitudinal health-behaviour research, such as: multi-sourced sampling, personalizing participant correspondence, adjusting survey characteristics and administration methods, sending pre-notification letters, and using larger incentives (Thornton et al., 2016; Subbaraman et al., 2015). In addition, advancements have also been made in evaluating numerous techniques for ensuring high follow-up rates (i.e. up to 99% at 1-year follow-up) including, community outreach, providing small study gifts, adjusting incentives based on promptness of survey completion, collecting collateral contact information, and maintaining consistent contact with participants between follow-up periods (Scott, 2004; Thomson et al., 2008; David

et al., 2013; Booker et al., 2011; Smith et al., 2017). Nevertheless, the effectiveness of some of these strategies appears to be limited (van Gelder et al., 2018), and many can be costly and resource intensive; along with being inappropriate for use in some samples. Likewise, many of these strategies do not address specific participant barriers to recruitment such as, overall lack of trust in research, misunderstanding the study purpose and/or participation requirements, and a lack of perceived individual benefits (Bonevski et al., 2014). As recruitment descriptions and study documents play a large role in participants' comprehension of the study and initial willingness to participate (e.g. information letters, consent forms, general correspondence), a greater understanding of how the language and readability of these documents impact participation rates and sample selection biases is warranted.

A growing body of literature has identified the need for study documents to encourage participation via plain language, however, the general trend is for researchers to use less comprehensible documents with higher reading levels (Ennis and Wykes, 2016; Foe and Larson, 2016; Friedman et al., 2014; Hamnes et al., 2016). Overall, the relationship between the language used in study material and participation rates has been understudied. While some researchers have provided evidence that optimally written study materials (e.g. shorter, "bulleted" lists, question/answer, lower reading levels) may slightly improve recruitment rates among depression (6.3% vs. 4.0%) and cardiovascular disease samples (24.0% vs. 21.9%), others have found no effect of such letters on the recruitment rates of cancer survivors (Hall et al., 2013; Man et al., 2015). Further, the authors are unaware of any research which has looked at this relationship among highly stigmatized samples, such as substance users, which may be more sensitive to language nuances.

This paper aims to further our knowledge in this area, by examining how one small text change in a recruitment document can impact recruitment and follow-up rates in a group of hazardous alcohol users interested in evaluating an online intervention for drinking. In a previously completed study (Cunningham et al., 2019), the authors employed a multi-stage recruitment process, such that participants were only enrolled into the study upon logging into a study portal with a pre-assigned password. The set of passwords used in the study, a randomly generated list of six letter words, contained what authors identified (post study-completion) as potentially inappropriate words for use within a substance using sample (e.g. winoes, nitwit) and a combination of high and low reading-level words (e.g. cymose, create). As all other parts of the study documents sent to participants were identical, this provided a unique opportunity to examine the impact of one small text change (i.e. using inappropriate vs. appropriate and low vs. high reading level passwords) on subsequent login rates (i.e. recruitment) and 6-month follow-up completions. As analyses were exploratory in nature, no a priori hypotheses guided the analyses.

2. Methods

2.1. Study design

As described elsewhere (Cunningham et al., 2019; Cunningham et al., 2017a), current drinkers were recruited from Amazon's Mechanical Turk (MTurk) in December 2016 to participate in an online randomized controlled trial (RCT) investigating the effectiveness of two Internet interventions for hazardous alcohol consumption. Recruitment into the study consisted of a multi-step process which included: (1) a brief screening survey (i.e. aged 18 and over, and weekly drinkers) and an online consent form; (2) a baseline survey to assess eligibility for the full RCT (Alcohol Use Identification Disorders Test (AUDIT) score ≥ 8 and consumption of at least 15 drinks per week); and (3) logging into the study portal with an assigned password which randomized participants into one of the two interventions or a control condition. All respondents who completed the baseline survey were paid $1.50 USD, regardless of whether they were eligible for the full RCT. In addition,

enrolled participants who completed the 6-month follow-up were compensated an additional $10 USD. Ethics approval for the research methods used in this study was provided by the standing ethics review committee of the Centre for Addiction and Mental Health (CAMH).

2.2. Thematic coding of passwords

Assigned passwords were randomly generated using an online Scrabble word dictionary. Passwords were dichotomously coded (0-No/ 1-Yes) across two thematic categories: (1) inappropriate, and (2) at the grade-8 reading level or below. Inappropriate passwords were defined as words whose common usage or definition was deemed offensive to the general public or unprofessional (e.g. profanity, sexually explicit, insulting, slang, religious). Examples of passwords coded as inappropriate were addict, brandy, cockup, nitwit, and sucker, in comparison to appropriate passwords which included ascend, fronts, prompt, tables, yellow. In addition, words phonetically similar sounding to an inappropriate word (e.g. uncock, testee) were coded as such. Grade-8 reading level and below was used to determine word recognition by the sample and categorize passwords into low and high readability based on previous literature suggesting that the average reading-grade level of substance users is between grade 8 and 10 (Johnson et al., 1995; Davis et al., 1993). Readability index scores (e.g. Flesch-Kincaid) were not computed as these types of scores are designed for groups of text and rely heavily on syllable counts. Instead, authors were presented with several grade-8 vocabulary lists (retrieved from online teaching resources) as a prime to guide coding. Examples of passwords coded at a grade-8 reading level or less included words like bushes, circle, legend, throat, and vendor, in comparison to passwords coded above the grade-8 reading level such as abomas, corymb, kevils, pusley and vivers. Passwords were independently coded by two of the authors, with good inter-rater agreement for both inappropriate password categorizations (92.5%) and grade-8 reading level or less (83.5%). Discrepancies between the two authors were resolved by a third author's coding of the passwords.

2.3. Data analysis

Bivariate analyses were performed to determine if password readability and inappropriate categorizations were randomly distributed across sample characteristics (see Table 1 for a list of variables). To test for sample representativeness across time, separate binomial logistic regressions were conducted to identify if (1) recruited participants differed from nonparticipants, and (2) 6-month follow-up completers differed from non-completers, across password assignment categorizations and key demographic and clinical characteristics at baseline. The main effect of password categorizations and sample characteristics, as well as, their interactions, were modelled. To determine which demographic and clinical characteristics should be entered into each respective logistic regression model, a series of chi-square tests and t-tests were performed for each sample to identify variables with a p value of 0.2 or less (Hosmer and Lemeshow, 1989; Zoran et al., 2008). A p-value of 0.2 was utilized as more traditional levels of 0.05 may fail to identify important variables (Zoran et al., 2008). A backwards, stepwise elimination process was used to remove variables, and interactions, from the logistic regression models if a p value of 0.1 or more was observed on the likelihood ratio test. This approach was used to reduce the problem of multi-collinearity and mitigate overfitting the model. However, as the intent of the study was to test for the effect of password categorization on recruitment and follow-up rates, password categorizations were forced into each logistic regression model, and are included in all final models. All other analyses were performed using SPSS, version 25.0.

Table 1

Bivariate comparisons within password categorizations for consenting participants.

| | Consenting participants N = 1002 | | | | | |
| | Inappropriate password | | | Grade-8 level password | | |
	No	Yes	*p*	No	Yes	*p*
Variables, %						
Male	58.1	60.3	0.636	58.6	58.1	0.894
Some post-secondary education	64.4	70.2	0.203	63.2	66.9	0.221
Married/common-law	42.6	39.7	0.545	42.4	42.1	0.915
Full/part time employed	68.3	76.0	0.085	67.1	71.3	0.146
Income < $20,000	20.9	20.7	0.955	23.0	18.8	0.098
High self deception[a]	48.9	45.5	0.474	49.2	47.9	0.679
High impression management[a]	50.4	46.3	0.396	51.9	48.1	0.230
Stage of change						
Pre-contemplation	39.0	37.2	0.350	38.3	39.3	0.854
Contemplation	41.8	47.9		43.4	41.7	
Action	19.2	14.9		18.3	19.0	
Ever received treatment	21.2	19.0	0.574	19.8	22.1	0.363
Study condition						
Control	32.9	28.9	0.379	31.3	33.5	0.447
Any intervention	67.1	71.1		68.7	66.5	
Variables, mean (SD)						
Age	35.0 (10.2)	33.9 (9.3)	0.293	34.7 (10.5)	34.9 (9.8)	0.731
AUDIT	17.9 (6.9)	18.3 (7.6)	0.590	17.9 (7.0)	18.0 (7.1)	0.717
# of weekly drinks	28.2 (15.1)	28.5 (17.3)	0.809	28.3 (16.0)	28.1 (14.7)	0.856

AUDIT – Alcohol Use Disorder Identification Test.

[a] Scale scores computed using the 40-item Balanced Inventory of Desirable Responding, high group was defined as those scoring above the median split (self-deception: 84; impression management: 72).

3. Results

3.1. Sample and study characteristics

In total, 3741 respondents completed the baseline survey, 1061 were found eligible to participate in the full RCT and invited to complete a 6-month follow-up, and 1002 agreed and were sent an email with a password to access the online study portal. In total, 511 logged-in and were recruited into the full RCT (Fig. 1). The latter two groups (n = 1002 and n = 511) comprise the samples used in the analyses examining factors associated with recruitment and 6-month follow-up, respectively.

The recruitment documents presented to participants were at various grade-level readabilities (i.e. Flesch-Kincaid) as computed by Microsoft Word: survey description (grade-9), consent form (grade-13), and invitation to login to study portal, not including the password (grade-8). Nearly half of the sample received a grade-8 level word (51.5%), however, only a small number of participants received an inappropriate password (12.1%). Further, the majority of passwords categorized as inappropriate were above the grade-8 reading level (62.0%). Bivariate analyses (Table 1) demonstrated no significant differences for inappropriate and grade-8 password allocation across demographic and clinical characteristics prior to enrolment (*p* > 0.05), resembling a random distribution of passwords.

3.2. Factors associated with recruitment

Initial chi-square analyses and *t*-tests within eligible participants (n = 1002) revealed five variables (gender, education, marital status, previous treatment, age) with a *p* value < 0.2, which were entered into the binomial logistic regression model along with inappropriate and
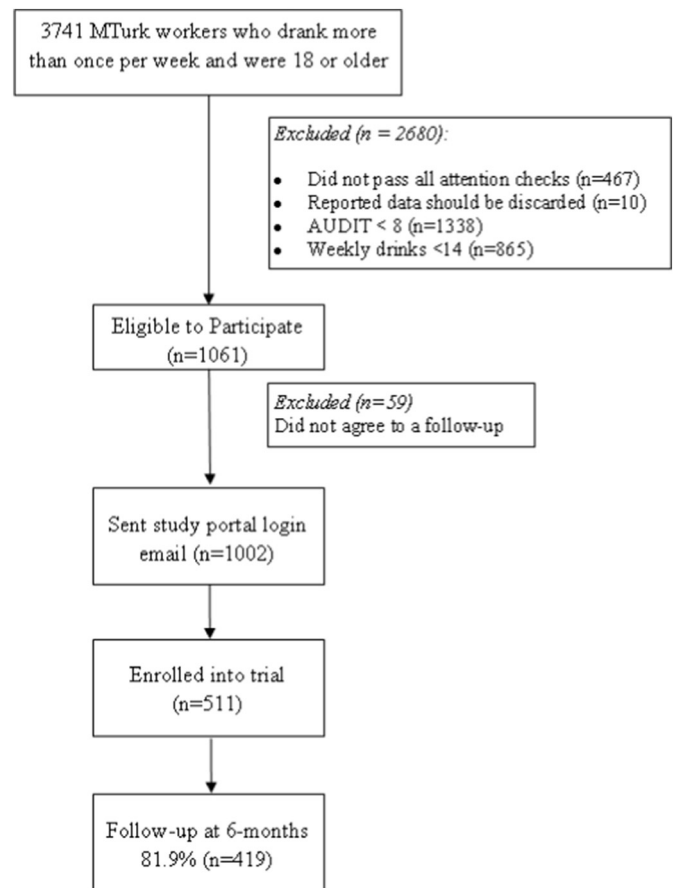

Fig. 1. Trial consort diagram.

grade-8 password categorizations, and interactions. The final model revealed recruitment was only significantly associated with marital status, age, and the inappropriate password by education interaction. In particular, being married/common-law and older significantly increased potential participants odds of enrolment into the study by 1.47 and 1.03 times, respectively. On the other hand, being assigned an inappropriate password had a different effect on enrolment depending on participants' level of education. While being assigned an inappropriate password increased the odds of enrolment for those without post-secondary education (1.82 times), this effect was not statistically significant (*p* = 0.111). Conversely, inappropriate passwords were a significant barrier to enrolment for those with a post-secondary education, and the odds of enrolment for these participants decreased by 26% (*p* = 0.044). Results and demographic characteristics are presented in Table 2.

3.3. Factors associated with follow-up at 6 months

Among the sample of recruited participants (n = 511), 419 (81.9%) were followed-up at 6-months. Overall, chi-square analyses and *t*-tests revealed four demographic characteristics which could potentially be associated with follow-up completion (*p* < 0.2). Education, income, self-deceptive enhancement (i.e. unconscious social desirable responding), and age, were entered into the binomial logistic regression model examining follow-up completion, along with inappropriate and grade-8 password categorizations, and interactions. The final model revealed that the odds of completing a follow-up were significantly increased for participants assigned a grade-8 password (1.95 times), with a reported income less than $20,000 (2.26 times), and scoring above the sample median on self-deceptive enhancement (1.99 times). No significant interactions between demographic variables and

Table 2
Results of chi-square, *t*-tests and binomial logistic regression analyses assessing baseline demographic, clinical, and password categorization variables associated with recruitment (N = 1002).

	Chi-square/*t*-test analysis				Binomial logistic regression	
	Drop-outs (n = 491)	Enrolled (n = 511)	Test statistic (df)	p[a]	Odds ratio (95% CI)	Likelihood ratio χ^2 (df), p[b]
Variables, n (%)						
Inappropriate password						
No	430 (87.6)	451 (88.3)	0.11 (1)	0.741[c,d]	1.82 (0.87–3.80)	0.60 (1), 0.111
Yes	61 (12.4)	60 (11.7)				
Grade-8 level password						
No	247 (50.3)	239 (46.8)	1.25 (1)	0.263[c,d]	1.18 (0.92–1.52)	0.17 (1), 0.202
Yes	244 (49.7)	272 (53.2)				
Male	298 (60.8)	286 (56.0)	2.42 (1)	0.120[c,e]		
Some post-secondary education	335 (68.2)	317 (62.0)	4.23 (1)	**0.040**[c]	0.83 (0.63–1.10)	−0.19 (1), 0.191
Married/common-law	179 (36.5)	244 (47.7)	13.09 (1)	**< 0.001**[c]	1.47 (1.14–1.91)	0.39 (1), **0.003**
Full/part time employed	338 (68.8)	356 (69.7)	0.08 (1)	0.776		
Income < $20,000	96 (19.6)	113 (22.1)	1.00 (1)	0.318		
High self deception	237 (48.3)	249 (48.7)	0.02 (1)	0.884		
High impression management	247 (50.3)	253 (49.5)	0.06 (1)	0.801		
Stage of change						
Pre-contemplation	189 (38.5)	200 (39.1)	0.35 (2)	0.841		
Contemplation	213 (43.4)	213 (41.7)				
Action	89 (18.1)	98 (19.2)				
Ever received treatment	93 (18.9)	117 (22.9)	2.37 (1)	0.124[c,e]		
Variables, mean (SD)[§]						
Age	33.4 (9.8)	36.3 (10.2)	−4.58	**< 0.001**[c]	1.03 (1.01–1.04)	0.03 (1), **< 0.001**
AUDIT	18.2 (7.0)	17.7 (7.0)	1.11	0.269		
# of weekly drinks	27.7 (14.8)	28.7 (15.8)	−1.00	0.318		
Interactions						
Inappropriate password × some post-secondary education	–	–	–	–	0.41 (0.17–0.98)	−0.90 (1), **0.044**

Bold denotes $p < 0.05$.

[§] Measured as a continuous variable.

[a] Variables with a p value of 0.2 or less were included in the initial binomial logistic regression as well as interactions with inappropriate and grade-8 level passwords, respectively. Only interactions included in the final model are presented.

[b] Variables with a p value of 0.1 or more on the likelihood ratio test were removed in a stepwise elimination process and only the final model is presented.

[c] Included in initial multiple regression analysis.

[d] To control for individuals assigned to different appropriateness and reading levels of passwords, these variables were included in all binomial logistic regression models despite p values > 0.2 on bivariate statistics and p value > 0.1 in likelihood ratio.

[e] Excluded from final model via a backward step-wise elimination process.

inappropriate or grade-8 passwords were observed. All demographic characteristics and results for this sample are presented in Table 3.

4. Discussion

This study provides preliminary evidence to suggest that one small text change in a study document can have significant implications for recruitment and follow-up rates, potentially resulting in sample selection biases within alcohol using samples. In particular, we found that assigning inappropriate passwords was a significant barrier for recruiting participants with at least some post-secondary education, as compared to those with a high school diploma or less. One possible explanation for this finding is that the use of inappropriate passwords jeopardized individuals' trust in the research or research team, or prompted participants to question personal benefits, perceived stigma, or the validity of the intervention; key barriers to participation in health research (Bonevski et al., 2014; Hunter et al., 2012). As those with a higher level of education may have been better able to recognize these passwords as offensive or unprofessional (62% of inappropriate words were above the grade-8 reading level), it is likely that decisions regarding continued participation for these individuals was impacted. Previous findings have suggested that individuals with a higher level of education are more likely to be undecided about participating in medical research generally (Trauth et al., 2000), thus receiving an inappropriate password may have prompted their withdrawal. Alternatively, those with less education may have been more likely to

overlook the inappropriate password and participate for personal benefits such as, financial incentives and/or treatment access.

Conversely, some demographic variables may facilitate enrolment. The results of this study support previous findings in the literature that older age groups and family relationships play a role in research participation (Trauth et al., 2000), as being older and in a common-law/married relationship was found to significantly increase participant odds of enrolling. Researchers should consider both factors that facilitate or hinder participation and follow-up completions when creating recruitment protocols and study documents for alcohol using samples to ensure that representative samples are recruited.

In addition, our findings illustrate how one mere word change in a recruitment document can have lasting effects for follow-up completions. That is, providing participants with a grade-8 password was significantly associated with greater odds of completing a follow-up. This is a particularly noteworthy finding as follow-up rates for this study were generally good (i.e. 82% at 6-months), making significant differences between completers and non-completers harder to detect. Although it is possible that grade-8 passwords may have encouraged greater participant interaction with the study, or improved study recall at follow-up, limited data regarding these factors precluded us from analyzing this further. Nonetheless, our finding was consistent with recent evidence suggesting that linguistic choices may indeed impact attitudes and perceptions of health-related research (Strekalova, 2018). A recent study by Strekalova (2018) found that defining a study as a "trial" deterred participation, while using the term "study" promoted

Table 3

Results of chi-square, *t*-tests and binomial logistic regression analyses assessing baseline demographic, clinical, and password categorization variables associated with completed follow-up (N = 511).

	Chi-square/*t*-test analysis				Binomial logistic regression	
	Lost to follow (n = 92)	Followed-up (n = 419)	Test statistic (df)	p^{a}	Odds ratio (95% CI)	Likelihood ratio χ^2 (df), p^{b}
Variables, n (%)						
Inappropriate password						
No	81 (88.0)	370 (88.3)	0.01 (1)	0.944[c,d]	0.91 (0.44–1.89)	− 0.09 (1), 0.807
Yes	11 (12.0)	49 (11.7)				
Grade-8 level password						
No	55 (59.8)	184 (43.9)	7.63 (1)	**0.006**[c]	1.95 (1.22–3.12)	0.67 (1), **0.005**
Yes	37 (40.2)	235 (56.1)				
Male	49 (53.3)	237 (56.6)	0.33 (1)	0.563		
Some post-secondary education	51 (55.4)	266 (63.5)	2.08 (1)	0.150[c]	1.53 (0.95–2.48)	0.43 (1), 0.082
Married/common-law	46 (50.0)	198 (47.3)	0.23 (1)	0.633		
Full/part time employed	62 (67.4)	294 (70.2)	0.28 (1)	0.600		
Income < $20,000	14 (15.2)	99 (23.6)	3.10 (1)	0.078[c]	2.26 (1.18–4.32)	0.81 (1), **0.014**
High self deception	33 (35.9)	216 (51.6)	7.43 (1)	**0.006**[c]	1.99 (1.23–3.22)	0.69 (1), **0.005**
High impression management	43 (46.7)	210 (50.1)	0.35 (1)	0.557		
Stage of change						
Pre-contemplation	36 (39.1)	164 (39.1)	0.01 (2)	0.994		
Contemplation	38 (41.3)	175 (41.8)				
Action	18 (19.6)	80 (19.1)				
Ever received treatment	23 (25.0)	94 (22.4)	0.28 (1)	0.596		
Study condition						
Control	30 (32.6)	148 (35.3)	0.25 (1)	0.621		
Any intervention	62 (67.4)	271 (64.7)				
Variables, mean (SD)[§]						
Age	34.3 (8.9)	36.7 (10.4)	− 2.30	**0.023**[c]	1.02 (1.00–1.05)	0.02 (1), 0.064
AUDIT	18.5 (7.1)	17.5 (7.0)	1.27	0.205		
# of weekly drinks	28.9 (16.6)	28.6 (15.7)	0.12	0.904		

AUDIT – Alcohol Use Disorders Identification Test. Bold denotes $p < 0.05$.

[§] Measured as a continuous variable.

[a] Variables with a *p* value of 0.2 or less were included in the initial binomial logistic regression as well as interactions with inappropriate and grade-8 level passwords, respectively. None of these interactions were included in the final model, therefore they are not presented.

[b] Variables with a *p* value of 0.1 or more on the likelihood ratio test were removed in a stepwise elimination process and only the final model is presented.

[c] Included in initial multiple regression analysis.

[d] To control for individuals assigned to different appropriateness levels of passwords, this variable was included in all binomial logistic regression models despite *p* value > 0.2 on bivariate statistics and *p* value > 0.1 in likelihood ratio.

greater participation rates among healthy individuals. Therefore, while readability and some linguistic choices may not be directly associated with enrolment rates as previous literature and this study demonstrate (Hall et al., 2013), it is possible for word choices in study descriptions to have long-term study effects on attrition rates; especially for vulnerable populations. While these findings are promising, further research is needed to better understand these relationships as well as associated ethical implications.

Lastly, our findings also demonstrated that those with lower incomes, and who scored higher on self-deception enhancement had greater odds of completing the 6-month survey. While is it likely that those with a lower income may have been highly motivated to complete the follow-up survey for the financial incentive (i.e. $10 USD), it is unknown why social desirable responding was associated with follow-up completions. Given the ongoing interest in the literature to examine social desirable responding effects on self-reported substance use (Perinelli and Gremigni, 2016; Latkin et al., 2017), our findings have implications for future research. It is recommended that investigators undertaking this type of research exercise caution when interpreting results if low follow-up rates are observed.

While this study provided an extreme example of how a small text change was associated with response rates and follow-up completions for alcohol survey research, there are some limitations which are important to consider. Firstly, this study was conducted using a sample recruited from Amazon's MTurk. Despite common concerns over the use of crowdsourcing platforms for substance-use research, alcohol-related data from this platform has been found to be of high-quality and previous studies conducted by the authors found MTurk workers to closely

approximate the drinking patterns and demographics of those recruited via other online or in-print sources (Cunningham et al., 2017a; Kim and Hodgins, 2017; Strickland and Stoops, 2018; Cunningham et al., 2017b). Secondly, analyses did not consider the possibility that some participants may not have enrolled into the study as a result of not receiving the invitation email from the study team. Third-party software, associated with MTurk, was used to email participants a link to the study portal and their assigned password, and thus it is possible that some emails were undeliverable or filtered as junk or spam. Thirdly, it should be noted that the readability of our consent form (grade 13) may have played a role in enrolment rates. While all participants received the same consent form, it may have shaped the relationship between education level, inappropriate passwords and enrolment, as well as subsequent follow-up completions. Lastly, it is important to acknowledge that passwords were qualitatively coded, and thus categorizations may have been inadvertently influenced by the personal experiences and education level of each coder. Future research should examine readability levels of entire documents using standardized measures when possible.

5. Conclusions

Overall, these findings highlight the need for researchers to (1) carefully scan study documents and processes for seemingly innocuous text which may unintentionally impact participant comprehension of the study, perceptions and willingness to participate, (2) consider linguistic choices and readability when developing study descriptions and recruitment documents and (3) exercise transparency in reporting

sample selection biases during recruitment and follow-up. Despite these recommendations, it is acknowledged that the results of this study may not be generalizable to all medical research. For example, interventions which require more invasive or complex procedures, rather than survey research, may face greater recruitment and retention barriers, which are not amenable to linguistic changes or readability scores.

The results of the study point to the need for researchers to carefully consider the language contained in recruitment documents when trying to ensure sample representativeness and low attrition rates. Study descriptions and research documents can have far greater implications on recruitment and follow-up rates for potential participants such as shaping public attitudes towards future research participation, and trust in research projects and researchers more generally. Future research should examine these associations in greater detail in substance-using samples, as much of the existent literature has focused on biomedical research.

List of abbreviations

AUDIT   Alcohol Use Disorders Identification Test
MTurk   Mechanical Turk
RCT   randomized controlled trial

Declarations

Ethics approval for the research methods used in this study was provided by the standing ethics review committee of the Centre for Addiction and Mental Health (CAMH).

Role of funding sources

Contributors

All authors have made an intellectual contribution to this research trial. JAC is the principal investigator, with overall responsibility for the project. All authors conceived the study and oversaw all aspects of the project, and were involved in the development of the protocol. AG wrote the first draft of this manuscript, and AG and CS conducted the data analyses. All authors have contributed to the manuscript drafting process, have read, and approved the final manuscript.

Data availability

The dataset supporting the conclusions of this article is available by reasonable request from the corresponding author.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

References

Barry, A.E., 2005. How attrition impacts the internal and external validity of longitudinal research. J. Sch. Health 75 (7), 267–270.

Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., et al., 2014. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med. Res. Methodol. 14 (42), 1–29.

Booker, C.L., Harding, S., Benzeval, M., 2011. A systematic review of the effect of retention methods in population-based cohort studies. BMC Public Health 11 (249), 1–12.

Capaldi, D., Patterson, G.R., 1987. An approach to the problem of recruitment and retention rates for longitudinal research. Behav. Assess. 9 (2), 19–177.

Cunningham, J.A., Godinho, A., Kushnir, V., 2017a. Using Mechanical Turk to recruit participants for Internet intervention research: experience from recruitment for four trials. BMC Med. Res. Methodol. 17 (1), 156.

Cunningham, J.A., Godinho, A., Kushnir, V., 2017b. Can Amazon's mechanical Turk be used to recruit participants for internet intervention trials? A pilot study involving a randomized control trial of a brief online intervention for hazardous alcohol use. Internet Interv. 10, 12–16.

Cunningham JA, Godinho A, Bertholet N. Outcomes of two randomized controlled trials, employing participants recruited through Mechanical Turk, of Internet interventions targeting unhealthy alcohol use. under review. 2019.

David, M.C., Alati, R., Ware, R.S., Kinner, S.A., 2013. Attrition in a longitudinal study with hard-to-reach participants was reduced by ongoing contact. J. Clin. Epidemiol. 66, 575–581.

Davis, T.C., Jackson, R.H., George, R.B., Long, S.W., Talley, D., Murphy, P.W., et al., 1993. Reading ability in patients in substance misuse treatment centers. International Journal of Addictions. 28, 571–582.

Ennis, L., Wykes, T., 2016. Sense and readability: participant information sheets for research studies. Br. J. Psychiatry 208 (2), 189–194.

Foe, G., Larson, E.L., 2016. Reading level and comprehension of research consent forms: an integrative review. Journal of Empirical Research on Human Research Ethics. 11 (1), 31–46.

Friedman, D.B., Kim, S., Tanner, A., Bergeron, C.D., Foster, C., General, K., 2014. How are we communicating about clinical trials? An assessment of the content and readability of recruitment resources. Contemporary Clinical Trials. 38, 275–283.

Hall, A.E., Sanson-Fisher, R.W., Lynagh, M.C., Threlfall, T., D'Este, C.A., 2013. Format and readability of an enhanced invitation letter did not affect participation rates in a cancer registry-based study: a randomized controlled trial. J. Clin. Epidemiol. 66, 85–94.

Hamnes, B., van Hustings-Eijk, Y., Primdahl, J., 2016. Readability of patient information and consent documents in rheumatological studies. BMC Medical Ethics. 17 (42), 1–9.

Hosmer, D., Lemeshow, S., 1989. Applied Logistic Regression. Wiley, New York, NY.

Hunter, J., Corcoran, K., Leeder, S., Phelps, K., 2012. Appealing to altruism is not enough: motivators for participating in health services research. Journal of Empirical Research on Human Research Ethics. 7 (3), 84–90.

Johnson, M.E., Fisher, D.G., Davis, D.C., Cagle, H.H., 1995. Reading abilities of drug users in Anchorage. Alaska. Journal of Drug Education. 25, 73–80.

Kim, H., Hodgins, D., 2017. Reliability and validity of data obtained from alcohol, cannabis, and gambling populations on Amazon's Mechanical Turk. Psychol. Addict. Behav. 31 (1), 85–94.

Kristman, V., Manno, M., Cote, P., 2004. Loss to follow-up in cohort studies: how much is too much? Eur. J. Epidemiol. 19, 751–760.

Latkin, C., Edwards, C., Davey-Rothwell, M., Tobin, K., 2017. The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore. Maryland. Addictive Behaviors. 73, 133–136.

Man, M.-S., Rick, J., Bower, P., 2015. Improving recruitment to a study of telehealth management for long-term condition in primary care: two embedded, randomised controlled trials of optimised patient information materials. Trials. 16, 309.

Perinelli, E., Gremigni, P., 2016. Use of social desirability scales in clinical psychology: a systematic review. J. Clin. Psychol. 72 (6), 534–551.

Scott, C.K., 2004. A replicable model for achieving over 90% follow-up rates in longitudinal studies of substance abusers. Drug Alcohol Depend. 74 (1), 21–36.

Smith, L.J., McNamara, P.J., King, A.C., 2017. Optimizing follow-up and study retention in the 21st century: advances from the front line in alcohol and tobacco research. Drug Alcohol Depend. 175, 171–178.

Strekalova, Y.A., 2018. Defining research: the effect of linguistic choices on the intentions to participate in clinical research. Clin. Nurs. Res. 27 (7), 790–799.

Strickland, J.C., Stoops, W.W., 2018. Feasibility, acceptability, and validity of crowd-sourcing for collecting longitudinal alcohol use data. J. Exp. Anal. Behav. 110 (1), 136–153.

Subbaraman, M.S., Lauder, A.B., Ritter, L.A., Stunz, A., Kaskutas, L.A., 2015. Multisource recruitment strategies for advancing addiction recovery research beyond treated samples. Journal of Community Psychology. 43 (5), 560–575.

Thomson, C.L., Thomson, C.L., Morley, K.C., Teesson, M., Sannibale, C., Haber, P.S., 2008. Issues with recruitment to randomised controlled trials in the drug and alcohol field: a literature review and Australian case study. Drug and Alcohol Review. 27 (2), 115–122.

Thornton, L.K., Harris, K., Baker, A.L., Johnson, M., Kay-Lambkin, F.J., 2016. Recruiting for addictions research via Facebook. Drug and Alcohol Review. 35, 494–502.

Trauth, J.M., Musa, D., Siminoff, K., Jewell, I.K., Ricci, E., 2000. Public attitudes regarding willingness to participate in medical research studies. Journal of Health and Social Policy. 12 (2), 23–43.

van Gelder, M.M.H.J., Vlenterie, R., IntHout, J., Engelen, L.J.L.P.G., Vrieling, A., van de Belt, T.H., 2018. Most response-inducing strategies do not increase participation in observational studies: a systematic review and meta-analysis. J. Clin. Epidemiol. 99, 1–13.

von Allmen, R.S., Weiss, S., Tevaearai, H.T., Kuemmerli, C., Tinner, C., Carrel, T.P., 2015. Completeness of follow-up determines validity of study findings: results of a prospective repeated measures cohort study. PLoS One 10 (10), e0140817.

Zoran, B., Gauss, C.H., Williams, D.K., Hosmer, D.W., 2008. Purposeful selection of variables in logistic regression. Source Code for Biology and Medicine. 3.