**Randomised trial seeking to induce the Hawthorne effect found no evidence for any effect on self-reported alcohol consumption online**

Jim McCambridge1,2 (jim.mccambridge@york.ac.uk)

Amanda Wilson1 (amanda.wilson@newcastle.edu.au)

John Attia1 (john.attia@newcastle.edu.au)

Natasha Weaver1 ([natasha.weaver@newcastle.edu.au](mailto:natasha.weaver@newcastle.edu.au))

Kypros Kypri1\* (kypros.kypri@newcastle.edu.au)

1Centre for Clinical Epidemiology and Biostatistics

School of Medicine and Public Health, University of Newcastle, Australia

2Department of Health Sciences, University of York, UK

\*Corresponding author and Guarantor:

Room 4104, HMRI Building, University of Newcastle, Callaghan NSW 2308 Australia  
+61 2 **4042 0536** (Fax +61 2 4042 0044)

**Role of the funding source:** This work was supported by the Australian Research Council (Discovery Grant DP1093809). The funder had no role in the study design or conduct or in the decision to publish the results.

**Abstract**

**Objective:** We tested the hypothesis that participants who know the behavioural focus of a study and are thus aware that a particular behaviour is being studied, will modify that behaviour, independently of any possible effect of assessment, thereby dismantling a Hawthorne effect into two putative components.

**Study Design and Setting:** We undertook a three-arm individually randomised trial online among students: Group A (control) were told they were completing a lifestyle survey; Group B were told the focus of the survey was alcohol consumption; Group C additionally answered 20 questions on their alcohol use and its consequences before answering the same lifestyle questions as Groups A and B. Non-drinkers were excluded and all groups were aware they would be followed-up after one month.

**Results:** Outcome data were obtained for 4583/5478 trial participants (84% follow-up rate). There were no differences between the three groups on primary (overall volume consumed) or secondary outcome measures (drinking frequency and amount per typical occasion) in the intervening four weeks.

**Conclusions:** There is no evidence that any form of Hawthorne effect exists in relation to self-reported alcohol consumption online among university students in usual research practice. Attention to study contexts is warranted for investigating research participation effects.

**Keywords**: Hawthorne effect; research participation effects; reactivity; alcohol; Internet; students

**Running Title**: Dismantling the Hawthorne effect online

**Word Count**: 2985

|  |
| --- |
| **What is new?**   * There is no evidence that any form of Hawthorne effect exists in relation to self-reported alcohol consumption online among university students in this large randomised trial. * Study findings are interpreted as suggesting that Hawthorne effect phenomena are likely to be unimportant in online alcohol studies. * There is a need to better understand the contexts in which research participation effects, which may threaten valid inferences, are more likely to occur. |

1. **Introduction**

Simply participating in a research study has the potential to modify behaviour in ways that can bias the findings, a phenomenon widely known as the ‘Hawthorne effect’ [1-3]. There is, however, little securely known about how any such effect works, or the circumstances in which it may arise [4]. Better understanding how participants react to what they are asked to do in research, and their experiences of the study conditions more broadly, may help achieve less biased study designs and inform the interpretation of existing findings [5, 6].

There is much uncertainty about what occurred in the original studies that inspired the naming of the Hawthorne effect, the various meanings attached to the term, and the resulting heterogeneity of objects of study in purposively designed evaluation studies of this phenomenon [4]. There are two principal types of studies of the Hawthorne effect, examining the effects of: 1) answering questions in interviews or by completing questionnaires; and 2) having behaviour directly observed or otherwise being made aware of having a particular behaviour studied [4]. These contrasting study types suggest two quite different mechanisms of effect: either that answering questions induces altered thinking about the behaviour, or that monitoring and surveillance in the context of research studies causes behaviour change; and it could be the case that both can operate simultaneously [4]. Conformity to perceived norms, or researcher expectations that may embody them, potentially links these two putative mechanisms, though a wide array of mechanisms have been proposed [7].

The effects of answering questions have been conceptualised in various ways in different topic areas and disciplines. Within social, health and consumer psychology, there is a large volume of research [8-12] on what is now known as the *question-behaviour effect*. Within the alcohol field, *assessment reactivity* has been the predominant conceptualisation of essentially the same phenomenon [13, 14]. An older and more recognisable construct known throughout the discipline of psychology is that of *demand characteristics* [15, 16], which incorporates both the ways in which questions are asked and other implicit features of research situations that participants may form expectations about what is being required of them. Demand characteristics are thus somewhat analogous to the two main operationalisations of the Hawthorne effect [17]. Investigations of demand characteristics belong to a rich and largely discontinued tradition within social psychology (due to changing norms about the acceptability of deception in non-laboratory research) of scrutinising unintended artefacts of research studies [18-20].

These considerations led us to develop the term *research participation effects* to refer broadly to the consequences arising from overlooked aspects of participating in research studies that may produce bias [6]. Although a rudimentary concept, it nonetheless connects a range of conventionally understood sources of bias (such as information and attrition bias) and provides a novel perspective that may guide dedicated studies which extend our understanding of the nature of bias [6].

There is evidence from randomised trials showing that asking people about their drinking without subsequent intervention can produce small reductions in self-reported alcohol consumption [14]. The present study is the third in a series of large online methodological experiments undertaken with university students in which alcohol consumption was measured to assess the outcome of a dedicated experimental manipulation. In the first we tested hypotheses about the effects of study design and allocation (ESDA), i.e., if knowing whether the study was a cohort (i.e., observational) study versus an intervention trial would affect university students’ subsequent drinking [21]. We expected that being told they were in an intervention trial might have a greater self-focusing--and thereby moderating--effect than being told they were in an observational study. We also investigated how the participants who thought they were in an intervention trial reacted to being told they had been randomised to the control versus intervention condition (allocation). We did not find support for either hypothesis, i.e., there were no significant differences in alcohol consumption between the three experimental groups when their drinking was assessed a month later [21].

Because we had to rely on self-report of outcome, we were concerned that any differences observed might merely reflect effects on reporting. Accordingly, in the second study, we sought to induce socially desirable reporting through asking about drinking in ways that might stigmatise participants, comparing this to a neutral presentation of questions as part of an assessment of health behaviour [22]. Again we found no differences between the groups and we concluded that in the web-based context, measurement of drinking is robust to the effects of question framing, at least within university student populations [22]. Note that self-report of drinking and other behaviours is nonetheless subject to a range of validity concerns other than framing effects [23, 24], and our focus is on usual research practice, not the possible effects of online surveillance. In the current study, using the same experimental paradigm, we investigated the effects of (1) participants knowing that the study’s focus was alcohol consumption in the absence of significant alcohol assessment; and (2) additionally completing alcohol assessments on their self-reported drinking behaviour one month later.

1. **Methods**
   1. *Design*

The overarching aims were to explore the content of those research participation effects commonly labelled as the Hawthorne effect. We undertook a three-arm individually randomised parallel groups trial (allocation ratio 1:1:1) to dismantle a hypothesised Hawthorne effect into putative assessment and awareness of observation components among university students in respect of their alcohol consumption. The existing literature is methodologically limited and thus complex to interpret [4], and the Hawthorne, Assessment Reactivity, Bias Investigational (HARBI) trial was prospectively registered with the Australian New Zealand Clinical Trials Registry (ACTRN12612000255886).

* 1. *Pilot work*

The trial was preceded by pilot research to refine the content of the experimental manipulations and study procedures. In the ESDA study investigating related research participation effects we found many participants spent little time with key web pages open [21]. We were concerned that some participants may not have read the text carefully, and thus were not fully exposed to the intended experimental manipulation.

For these reasons, the baseline instrument underwent two rounds of pilot testing at an Australian university. Students on campus were invited by a trained interviewer (AW) to assist with the design of a health behaviour survey using methods we developed earlier [25]. Those providing informed consent completed the questionnaire on a computer, taking approximately five minutes to complete it., and then took part in a face-to-face interview for a further five minutes. In the first round of piloting two thirds of participants (20/29) indicated they were unlikely to carefully read the introductory information in such a survey, which would have undermined our intended experimental manipulation. We therefore added a mandatory manipulation check to the study procedures, confirming that the relevant experimental groups knew this study was about alcohol, involved repeated measurement, and was applicable to non-drinkers. Correct information was repeated in the event of incorrect answers to the questions presented to participants, which they had to complete before progressing to the next webpage. This was successful, with the second round of piloting yielding 100% (24/24) awareness that participation involved repeated surveys of alcohol consumption.

*2.3. Participants*

We conducted the trial in four New Zealand universities. All students (approximate n of 60,000) without any restriction by age or enrolment type, were invited to complete a web survey in a message sent to their student e-mail address by the university administration. The message contained a hyperlink to a Participant Information Sheet which invited participation in a brief student lifestyle study and presented a button to indicate consent and start the survey. Two alcohol questions were included within the student lifestyle questionnaire and used to identify non-drinkers who completed the research procedures but were not eligible for participation in the trial. All trial participants thus reported consuming alcohol in the previous 4 weeks.

* 1. *Interventions/Experimental manipulations*

The detailed content of the interventions i.e., the text that comprised the experimental manipulation is presented in Figure 1. Briefly, Group A (the control group) was asked to complete a questionnaire on “Student Lifestyles”, Group B was advised that the study focus was their alcohol use, and Group C was also asked 20 questions about their drinking. All three groups were advised they would be followed up again in one month.

All surveys took less than 10 minutes to complete.

* 1. *Outcomes*

The pre-specified primary outcome was self-reported volume of alcohol consumption in the previous 4 weeks. This was computed as the product of the two secondary outcomes: frequency of days drinking and quantity (number of standard drinks) per drinking day, which were asked as follows: 1) On how many days in the last 4 weeks did you have a drink containing alcohol? 2) How many Standard Drinks containing alcohol did you have on a typical day when you were drinking in the last 4 weeks? Both questions were embedded within the lifestyle survey administered at baseline.

* 1. *Sample size*

The study was designed to detect between-group differences in effect size of d=0.1. This small effect size was chosen following partitioning of previously observed effect sizes in a systematic review of alcohol assessment reactivity [14]. To achieve power of 0.8, with an alpha 0.05 and a two-sided test required 1946 participants per group (5838 in total).

* 1. *Randomisation*

Randomisation without any restrictions was implemented using a computerised random number generator. All study procedures were computerised and automated ensuring that the researchers were blind to randomization, allocation, and assessment of outcome, and there was no direct contact with participants that could have compromised randomisation.

* 1. *Blinding*

All three randomised groups were blinded to the true purpose of the study, as this would have interfered with hypothesis testing. We have offered ethical justifications for the use of deception in such studies elsewhere [26] and make clear that the University of Otago’s Human Research Ethics Committee approved the conduct of this study. In order to conceal the alcohol focus of the study, no attempt was made to exclude non-drinkers from online participation in the surveys though they were not trial participants. All participants were given to understand that this was a brief student lifestyle study involving the completion of two surveys 4 weeks apart. No participant was aware of randomisation.

* 1. *Statistical Methods*

We used linear regression models with repeated measures fitted by Generalised Estimating Equations for each of the three outcomes, testing for differences in outcomes between each pair of groups and across all three groups.

1. **Results**

The CONSORT flowchart in Figure 2 shows that 6909 students (both drinkers and non-drinkers) initially consented, including n=5478 drinkers within the previous 4 weeks who were eligible for trial participation and were randomised to one of the three experimental groups. Trial enrolment took place in August-September 2012 and follow-up one month later. Table 1 shows that randomisation worked well and attrition was unproblematic as the three experimental groups had similar distributions of demographic characteristics.

Table 2 summarises alcohol consumption at baseline and follow-up by experimental group. This shows all three groups were similar in drinking at baseline, and that there were no differences in outcomes, i.e., the groups were not found to change differently over time. None of the pairwise comparisons showed a significant difference at follow-up. In contrast to the pilot study, Table 3 indicates that the experimental manipulation was similarly unsuccessful for Groups B and C in approximately 24% of participants.

1. **Discussion**

This study found no evidence that any form of Hawthorne effect exists in relation to self-reported alcohol consumption online among university students. Previous findings on assessment of this behaviour in this population online are mixed; some randomised trials have identified small effects for some outcomes [27-29], whereas others find no such effects [30, 31]. An earlier systematic review of randomised studies of alcohol assessment found some evidence of small effects and concluded that answering questions “appears to exert a subtle influence on subsequent self-reported drinking behaviour among students” [14], partly as there were no effects in randomised studies in non-student populations. It is noteworthy that in this earlier review [14] there were effects observed in trials investigating self-completion of pen-and-paper questionnaires [32] and interview administration [33] among students respectively. There are no prior randomised trials of simply being aware of the possible effects of a study focus on alcohol as investigated here.

Beyond alcohol, the larger literature on question-behaviour effects provides evidence of small effects on both self-reported and objectively ascertained outcomes, with much unexplained heterogeneity among studies, in a literature vulnerable to bias [8, 10, 12]. French and Sutton [34] identify larger assessment effects on emotion and cognition than on behaviour in a narrative review. Some meta-analytic reviews are restricted to trials [8], whereas others also synthesise non-randomised studies [9], with the latter identifying much larger effects than the former. These meta-analytic reviews identify larger effect sizes among students, and do not generally suggest that alcohol is clearly different than other behaviours, though effects are smaller with more challenging behaviours to perform. Effects are larger in promoting socially desirable behaviours than in reducing socially undesirable behaviours [9], as was the nature of the original Hawthorne factory research artefact. Wilding and colleagues [9] identify larger effects in laboratory settings and with face-to-face questioning, with online questions similar to mail and phone. These studies were largely undertaken by psychologists, and there has been a strong emphasis on question types and content [35], as well as a search for putative mechanisms underlying effects, which has proven difficult to progress [11]. The prior systematic review of the Hawthorne effect [4] also found the heterogeneity in previous studies of this construct challenging to interpret, as did another systematic review of Solomon 4-group designs [36].

The existing literature arguably requires much greater attention to both conceptualisation and empirical investigation of the effects of study contexts [37]. For example, it has been shown that completion of the same alcohol questionnaires in different contexts will produce different findings [38]. It has previously been suggested that question-behaviour effect studies need to pay greater attention to study contexts [39], and we suggest this is a particularly fruitful direction for studies of Hawthorne effect phenomena more generally. Indeed, study context is what shapes the original emphasis on observation, monitoring and surveillance [1-3, 40]. We have previously suggested that how participants interact with what they are asked to do in the specific contexts of research studies, i.e. participant engagement with study context, may be a fruitful line of enquiry [6]. Qualitative studies already suggest that there is much value in asking participants about their experiences in research to better understand the data they provide, and the potential implications for bias [41, 42].

The interpretation of this study’s findings needs to address study limitations. We fell a little below the target sample size, and suggest that reaching it would not have altered the study findings. We originally were attracted to the online university context as offering a pragmatic opportunity to recruit large numbers of research participants. An inherent limitation of this context is reliance on self-report of drinking behaviour [24]. Although in treatment contexts the conclusion has been reached that self-report of drinking is valid [43], this is not clear in other settings [44], although computerised assessments of sensitive behaviours may be more valid [45]. Note also that invalid self-report may itself be conceptualised as a research participation effect [6].

Another substantial study limitation lies in the lack of optimisation of the experimental manipulations, where despite careful pilot work, approximately one quarter of those randomised to the experimental groups, who were all told that the study focus was alcohol, did not confirm this approximately 5 minutes later. Interestingly, the asking of 20 alcohol questions did not make any difference in the randomised comparison to the simple study focus manipulation, and it may be that the asking of 18 non-alcohol-related questions was responsible.

The earlier literature on assessment reactivity in alcohol treatment contexts where assessments were typically lengthy did consider that the overall burden of assessment was likely to be important to the stimulation of unintended effects [13, 46, 47]. It is possible that a stronger contrast between Groups B and C, achieved by asking many more questions may show some effects, though this would be untypical of how online research in this area is conducted, where participant burden is a consideration in relation to preventing attrition [48].

It also appears that the online context is not conducive to the elicitation of a Hawthorne effect sense of being observed, perhaps because of the ways that participants engage with study and non-study requests for information and/or the nature of the medium. This suggests a much stronger surveillance manipulation would be needed to evoke a sense of being observed in brief contacts online. This would be challenging to investigate, for example, without compromising study retention. It may be that, despite the capacity for access to large numbers of participants, the online study context imposes limitations to further investigations of surveillance or that a different kind of manipulation is needed than was evaluated here. The effects of online surveillance could become the focus of social psychology investigations, though our primary focus is usual practices in research. Strengths of this study include security against any compromise of randomisation due to the automated nature of the study.

We conclude on the basis of the evidence examined here that there is no particular need for concern about Hawthorne type phenomena in online alcohol studies among university students in the way they are usually conducted, and perhaps also in other online populations. On the basis of our studies, we suggest that future research in this area may be most usefully advanced in research studies in which there is direct person-to-person contact which evokes social desirability concerns. Future studies of research participation effects should also identify other features of contexts in which there is a basis for concern about such effects, be explicit about the underlying reasoning to further develop appropriate concepts, and use high quality study designs to identify where these issues pose threats to valid inference.

**Conflicts of interest**

None.

**Authors’ contributions**

JM and KK conceived of and designed the trial and together with JA obtained the funding to undertake it. AW conducted the pilot study, and developed and implemented the trial procedures under the direction of KK. NW analysed the data. JM led the drafting of the manuscript with inputs from KK. All authors approved the final manuscript for submission.

**Acknowledgements**

We are grateful to the participating universities.

**Table 1. Demographic characteristics of the experimental groups**

|  | | **Experimental Group** | | |  |
| --- | --- | --- | --- | --- | --- |
|  |  | A (n=1797) | B (n=1853) | C (n=1828) | All (N=5478) | |
| Age, mean (SD) | | 24.0 (7.9) | 24.2 (8.1) | 24.2 (8.1) | 24.1 (8.1) | |
| Gender |  |  |  |  |  | |
|  | Men | 624 (35%) | 622 (34%) | 653 (36%) | 1899  (35%) | |
|  | Women | 1173 (65%) | 1231 (66%) | 1175 (64%) | 3579  (65%) | |
| Living arrangements | |  |  |  |  | |
|  | Share a rented flat or house | 966 (54%) | 998 (54%) | 992 (54%) | 2956  (54%) | |
|  | Living in a residential college or hall | 228 (13%) | 252 (14%) | 236 (13%) | 716  (13%) | |
|  | Living with parent(s) or guardian(s) | 238 (13%) | 228 (12%) | 219 (12%) | 685  (13%) | |
|  | Living in my own home | 333 (19%) | 327 (18%) | 342 (19%) | 1002  (18%) | |
|  | Boarding | 20 (1.1%) | 29 (1.6%) | 22 (1.2%) | 71 (1.3%) | |
|  | Living elsewhere | 8 (0.4%) | 7 (0.4%) | 13 (0.7%) | 28 (0.5%) | |
|  | Prefer not to answer | 4 (0.2%) | 12 (0.6%) | 4 (0.2%) | 20 (0.4%) | |

**Table 2. Alcohol consumption at baseline and follow-up by experimental group**

|  |  |  | | **Experimental Group** | | |  | **P-value** | |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Outcome** | Time |  | Statistic | A (n=1485) | B (n=1532) | C (n=1566) | Overall | B vs. A | C vs. A | B vs. C |
| Total volume consumed | Baseline |  | mean (SD) | 24.6 (31.8) | 23.8 (31.8) | 23.0 (30.6) |  |  |  |  |
|  |  |  | median  (min, max) | 15 (1, 336) | 14 (1, 459) | 13 (1, 476) |  |  |  |  |
|  | Follow-up |  | mean (SD) | 18.4 (23.3) | 17.5 (23.8) | 17.5 (23.2) | 0.66 | 0.89 | 0.51 | 0.39 |
|  |  |  | median  (min, max) | 12 (0, 240) | 10 (0, 360) | 10 (0, 308) |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Days drinking alcohol | Baseline |  | mean (SD) | 5.9 (5.2) | 5.8 (4.9) | 5.5 (5.0) |  |  |  |  |
|  |  |  | median  (min, max) | 4 (1, 28) | 4 (1, 28) | 4 (1, 28) |  |  |  |  |
|  | Follow-up |  | mean (SD) | 5.2 (5.5) | 5.0 (5.1) | 4.9 (5.0) | 0.26 | 0.22 | 0.80 | 0.12 |
|  |  |  | median  (min, max) | 4 (0, 28) | 4 (0, 28) | 4 (0, 28) |  |  |  |  |
| Drinks per drinking day | Baseline |  | mean (SD) | 4.3 (3.6) | 4.2 (3.6) | 4.2 (3.6) |  |  |  |  |
|  |  |  | median  (min, max) | 3 (1, 30) | 3 (1, 27) | 3 (1, 22) |  |  |  |  |
|  | Follow-up |  | mean (SD) | 3.5 (3.3) | 3.4 (3.4) | 3.4 (3.3) | 0.87 | 0.72 | 0.86 | 0.60 |
|  |  |  | median  (min, max) | 2 (0, 24) | 2 (0, 30) | 2 (0, 30) |  |  |  |  |

**Table 3. Experimental manipulation checks for Hawthorne effect groups**

|  | | **Experimental Group** | | |
| --- | --- | --- | --- | --- |
| **Question** | **Class** | **B (n=1853)** | **C (n=1828)** | **P-value** |
| Study purpose | Your drinking [correct] | 1418 (77%) | 1372 (75%) | 0.2035 |
|  | Health behaviour [incorrect] | 433 (23%) | 456 (25%) |  |
|  | Other [incorrect] | 2 (0.1%) |  |  |
| Study requirements | One now and another in 4 weeks [correct] | 1779 (96%) | 1741 (95%) | 0.0424 |
|  | Just one now [incorrect] | 59 (3.2%) | 80 (4.4%) |  |
|  | Other [incorrect] | 15 (0.8%) | 7 (0.4%) |  |
| Study eligibility | All students [correct] | 1804 (97%) | 1773 (97%) | 0.7826 |
|  | Only students who drink alcohol [incorrect] | 43 (2.3%) | 49 (2.7%) |  |
|  | Other [incorrect] | 6 (0.3%) | 6 (0.3%) |  |

University students invited by e-mail to participate in a “**Brief Student Lifestyle Study**”

Consented and randomised

**A**

**Control**

Webpage header: Student Lifestyles

Page 1: Student Lifestyles Questionnaire:

Manipulation check confirming that group A participants knew this study was about lifestyles and was applicable to non-drinkers.

The following questions concern student lifestyle. For each activity below, please indicate (1) on how many days you did it and (2) how much per day, in the last 4 weeks.

* Watching television
* Using the Internet for entertainment
* Studying
* Drinking alcohol
* Going out or socialising (e.g., going to restaurants, the cinema, concerts or bands)
* Playing sport or exercising
* Being involved in political activity
* Other (please specify, e.g., spending time with family)

Page 2: Thank you for completing the survey. We will e-mail you again in one month with a link to a final, short survey.

**C**

**Assessment of behaviour**

Webpage header: Student Lifestyles

Page 1: This study is about your **alcohol** use, how it relates to your lifestyle, and how it changes over time.Even if you do not drink alcohol please continue with the survey. We will measure your drinking now and in the follow-up survey.

*Manipulation check (see Table 3)*

Page 2: Alcohol assessment questions:

AUDIT items 4-10;

AREAS scale (5 items)

List 3 things you like about your drinking:

List 3 things that you don’t like about your drinking:

Page 3: Thank you for completing the baseline survey. We will e-mail you again in one month with a link to a final, short survey.

**B**

**Behavioural focus highlighted**

Webpage header: Student Lifestyles

Page 1: This study is about your **alcohol** use, how it relates to your lifestyle, and how it changes over time.Even if you do not drink alcohol please continue with the survey. We will measure your drinking now and in the follow-up survey.

*Manipulation check (see Table 3)*

Page 2: Student Lifestyles Questionnaire (as in Group A)

Page 3: Thank you for completing the baseline survey. We will e-mail you again in one month with a link to a final, short survey.

**1-month follow-up**

Webpage header: Student Lifestyles Survey 2

Page 1: What would you say was the design of the study?

Please select one of the following: (a) Two surveys of student lifestyle, (b) Following students’ drinking over time, (c)

Assessing students’ drinking over time

Thank you for participating in this study. If you would like to enter the draw for an iPad, please provide your e-mail address below.

**Figure 1. Study design and experimental conditions**

Approximately 60,000 university students invited by e-mail to participate

6909 consented and randomised

Group B

1853/2294 eligible and exposed to experimental manipulation

Group C

1828/2276 eligible and exposed to experimental manipulation

Group A

1797/2339 eligible and exposed to experimental manipulation

1566 (86%) assessed at follow-up and included in analyses

1532 (83%) assessed at follow-up and included in analyses

1485 (83%) assessed at follow-up and included in analyses

**Figure 2. CONSORT diagram**

**References**

1. French, J.R.P., *Experiments in field settings*, in *Research methods in the behavioral sciences*, L. Festinger, Katz, D., Editor. 1953, Holt, Rinehart & Winston: New York.

2. Parsons, H.M., *What happened at Hawthorne?* Science, 1974. **181**: p. 922-932.

3. Gillespie, R., *Manufacturing knowledge: a history of the Hawthorne experiments*. 1991, Cambridge: Cambridge University Press.

4. McCambridge, J., J. Witton, and D. Elbourne, *Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects.* J Clin Epidemiol, 2014. **67**(3): p. 267-277.

5. McCambridge, J., K. Kypri, and D. Elbourne, *In randomisation we trust? There are overlooked problems in experimenting with people in behavioural intervention trials.* J Clin Epidemiol, 2014. **67**(3): p. 247-253.

6. McCambridge, J., K. Kypri, and D. Elbourne, *Research participation effects: A skeleton in the methodological cupboard* J Clin Epidemiol, 2014. **67**: p. 845-849.

7. Chiesa, M. and S. Hobbs, *Making sense of social research: How useful is the Hawthorne Effect?* European Journal of Social Psychology, 2008. **38**(1): p. 67-74.

8. Rodrigues, A.M., et al., *The Question-Behavior Effect: Genuine Effect or Spurious Phenomenon? A Systematic Review of Randomized Controlled Trials With Meta-Analyses.* Health Psychology, 2015. **34**(1): p. 61-78.

9. Wilding, S., et al., *The question-behaviour effect: A theoretical and methodological review and meta-analysis.* European Review of Social Psychology, 2016. **27**(1): p. 196-230.

10. Wood, C., et al., *The Impact of Asking Intention or Self-Prediction Questions on Subsequent Behavior: A Meta-Analysis.* Personality and Social Psychology Review, 2016. **20**(3): p. 245-268.

11. Spangenberg, E.R., et al., *A meta-analytic synthesis of the question-behavior effect.* Journal of Consumer Psychology, 2016. **26**(3): p. 441-458.

12. Rodrigues, A.M., D.P. French, and F.F. Sniehotta, *Commentary: The Impact of Asking Intention or Self-Prediction Questions on Subsequent Behavior: A Meta-Analysis.* Frontiers in Psychology, 2016. **7**.

13. McCambridge, J., *Research assessments: instruments of bias and brief interventions of the future?* Addiction, 2009. **104**(8): p. 1311-2.

14. McCambridge, J. and K. Kypri, *Can simply answering research questions change behaviour? Systematic review and meta analyses of brief alcohol intervention trials.* PLoS One, 2011. **6**(10): p. e23748.

15. Orne, M.T., *The nature of hypnosis: Artifact and essence.* J. Abnorm. Soc. Psychol, 1959. **58**: p. 277-299.

16. Orne, M.T., *On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications.* American Psychologist, 1962. **17**: p. 776-783.

17. McCambridge, J., M. de Bruin, and J. Witton, *The effects of demand characteristics on research participant behaviours in non-laboratory settings: A systematic review.* PLoS ONE, 2012. **7**(6): p. e39116.

18. Webb, E.J., Campbell, D.T., Schwartz, R.D., Sechrest, L., *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. 1966, Oxford: Rand McNally.

19. Rosnow, R.L., Rosenthal, R. , *People studying people: Artifacts and ethics in behavioral research*. 1997, New York: Freeman.

20. Sharpe, D. and W.J. Whelton, *Frightened by an Old Scarecrow: The Remarkable Resilience of Demand Characteristics.* Review of General Psychology, 2016. **20**(4): p. 349-368.

21. Kypri, K., et al., *Effects of study design and allocation on self-reported alcohol consumption: randomized trial.* Trials, 2015. **16**: p. 127.

22. Kypri, K., et al., *Social Desirability Bias in the Reporting of Alcohol Consumption: A Randomized Trial.* J Stud Alcohol Drugs, 2016. **77**(3): p. 526-31.

23. Tourangeau, R. and T. Yan, *Sensitive questions in surveys.* Psychol Bull, 2007. **133**(5): p. 859-83.

24. Davis, C.G., J. Thake, and N. Vilhena, *Social desirability biases in self-reported alcohol consumption and harms.* Addict Behav, 2010. **35**(4): p. 302-11.

25. Hallett, J., et al., *Development of a Web-based alcohol intervention for university students: processes and challenges.* Drug Alcohol Rev, 2009. **28**(1): p. 31-9.

26. McCambridge J, K., K., Bendtsen, P., Porter, J., *The use of deception in public health behavioural intervention trials: a case study of three online alcohol trials* Am J Bioeth, 2013. **13**(11): p. 39-47.

27. Kypri, K., et al., *Assessment may conceal therapeutic benefit: findings from a randomized controlled trial for hazardous drinking.* Addiction, 2007. **102**(1): p. 62-70.

28. McCambridge, J., et al., *Alcohol assessment and feedback by e-mail for university students: main findings from randomised controlled trial.* Br J Psychiatry, 2013. **203**: p. 334-340.

29. Walters, S.T., et al., *Reactivity to alcohol assessment measures: an experimental test.* Addiction, 2009. **104**(8): p. 1305-10.

30. Bendtsen, P., McCambridge. J., Bendtsen, M., Karlsson, N., Nilsen, P. , *RCT of the effectiveness of proactive mail based alcohol Internet intervention with university students: dismantling the assessment and feedback components.* J Med Internet Res 2012. **14**(5): p. e142.

31. Kypri, K. and H.M. McAnally, *Randomized controlled trial of a web-based primary care intervention for multiple health risk behaviors.* Prev Med, 2005. **41**(3-4): p. 761-6.

32. McCambridge, J., Day, M., *Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking.* Addiction, 2008. **103**: p. 241-248.

33. Carey, K.B., et al., *Brief motivational interventions for heavy college drinkers: A randomized controlled trial.* Journal of Consulting and Clinical Psychology, 2006. **74**(5): p. 943 - 954.

34. French, D.P. and S. Sutton, *Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it?* Br J Health Psychol, 2010. **15**(Pt 3): p. 453-68.

35. Mankarious, E. and E. Kothe, *A meta-analysis of the effects of measuring theory of planned behaviour constructs on behaviour within prospective studies.* Health Psychology Review, 2015. **9**(2): p. 190-204.

36. McCambridge, J., et al., *Can research assessments themselves cause bias in behaviour change trials? A systematic review of evidence from Solomon 4-group studies.* PLoS ONE, 2011. **6**(10): p. e25223.

37. McCambridge, J., *From question-behaviour effects in trials to the social psychology of research participation.* Psychol Health, 2015. **30**(1): p. 72-84.

38. Cooke, R. and D.P. French, *The role of context and timeframe in moderating relationships within the theory of planned behaviour.* Psychol Health, 2011. **26**(9): p. 1225-40.

39. Sprott, D.E., Spangenberg, E.R., Block, L.G., Fitzsimons, G.J., Moritz, V.G., Williams, P. , *The question-behavior effect: What we know and where we go from here.* Social Influence, 2006. **1**(2): p. 128-137.

40. Gale, E.A.M., *The Hawthorne studies – a fable for our times?* Quarterly Journal of Medicine, 2004. **97**: p. 439-449.

41. Locock, L. and L. Smith, *Personal experiences of taking part in clinical trials - a qualitative study.* Patient Educ Couns, 2011. **84**(3): p. 303-9.

42. McCambridge, J., et al., *Patient preferences and performance bias in a weight loss trial with a usual care arm.* Patient Educ Couns, 2014. **95**(2): p. 243-247.

43. Babor, T.F., et al., *Talk is cheap: measuring drinking outcomes in clinical trials.* J Stud Alcohol, 2000. **61**(1): p. 55-63.

44. Noknoy, S., et al., *RCT of effectiveness of motivational enhancement therapy delivered by nurses for hazardous drinkers in primary care units in Thailand.* Alcohol Alcohol, 2010. **45**(3): p. 263-70.

45. Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Pleck, J.H., Sonenstein F.L., *Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting with Computer Survey Technology* Science, 1998. **280**: p. 867-873

46. Gastfriend, D.R., et al., *Developing a baseline assessment battery: balancing patient time burden with essential clinical and research monitoring.* J Stud Alcohol Suppl, 2005(15): p. 94-103; discussion 92-3.

47. Orford, J., et al., *The clients' perspective on change during treatment for an alcohol problem: qualitative analysis of follow-up interviews in the UK Alcohol Treatment Trial.* Addiction, 2006. **101**(1): p. 60-8.

48. McCambridge, J., et al., *Impact of length or relevance of questionnaires on attrition in online trials: randomized controlled trial.* J Med Internet Res, 2011. **13**(4): p. e96.