



METHOD ARTICLE

REVISED Updating the evidence on the effectiveness of the alcohol reduction app, Drink Less: using Bayes factors to analyse trial datasets supplemented with extended recruitment [version 2; peer review: 2 approved]

Claire Garnett ¹, Susan Michie², Robert West ¹, Jamie Brown¹

¹Department of Behavioural Science and Health, University College London, London, UK

²Department of Clinical, Education and Health Psychology, University College London, London, UK

v2 First published: 29 Jan 2019, 8:114 (<https://doi.org/10.12688/f1000research.17952.1>)

Latest published: 15 Jul 2019, 8:114 (<https://doi.org/10.12688/f1000research.17952.2>)

Abstract

Background: A factorial experiment evaluating the Drink Less app found no clear evidence for main effects of enhanced versus minimal versions of five components but some evidence for an interaction effect. Bayes factors (BFs) showed the data to be insensitive. This study examined the use of BFs to update the evidence with further recruitment.

Methods: A between-subject factorial experiment evaluated the main and two-way interaction effects of enhanced versus minimal version of five components of Drink Less. Participants were excessive drinkers, aged 18+, and living in the UK. After the required sample size was reached (n=672), additional data were collected for five months. Outcome measures were change in past week alcohol consumption and Alcohol Use Disorders Identification Test (AUDIT) score at one-month follow-up, amongst responders only (those who completed the questionnaire). BFs (with a half-normal distribution) were calculated (BF<0.33 indicate evidence for null hypothesis; 0.33<BF<3 indicate data are insensitive).

Results: Of the sample of 2586, 342 (13.2%) responded to follow-up. Data were mainly insensitive but tended to support there being no large main effects of the enhanced version of individual components on consumption (0.22<BF<0.83) or AUDIT score (0.14<BF<0.98). Data no longer supported there being two-way interaction effects (0.31<BF<1.99). In an additional exploratory analysis, participants receiving four of the components averaged a numerically greater reduction in consumption than those not receiving any (21.6 versus 12.1 units), but the data were insensitive (BF=1.42).

Conclusions: Data from extended recruitment in a factorial experiment evaluating components of Drink Less remained insensitive but tended towards individual and pairs of components not having a large effect. In an exploratory analysis, there was weak, anecdotal evidence for a synergistic effect of four components. In the event of uncertain results, calculating BFs can be used to update the strength of evidence of a dataset supplemented with extended recruitment.

Open Peer Review

Reviewer Status

	Invited Reviewers	
	1	2
REVISED		
version 2	report	report
published 15 Jul 2019		
version 1		
published 29 Jan 2019	report	report

1 **Daniel Leightley** , King's College London, London, UK

2 **Nick Heather** , Northumbria University, Newcastle upon Tyne, UK

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

Bayes Factors, digital interventions, alcohol reduction, smartphone apps



This article is included in the **University College London** collection.

Corresponding author: Claire Garnett (c.garnett@ucl.ac.uk)

Author roles: **Garnett C:** Conceptualization, Data Curation, Formal Analysis, Investigation, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Michie S:** Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing; **West R:** Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing; **Brown J:** Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: JB and RW are funded by Cancer Research UK (CRUK; C1417/A22962). CG and SM are funded by CRUK and the NIHR School for Public Health Research (SPHR). Drink Less was funded by the NIHR SPHR, the UK Centre for Tobacco and Alcohol Studies (UKCTAS), the Society for the Study of Addiction (SSA), and CRUK. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The research team is part of the UKCTAS, a UKCRC Public Health Research Centre of Excellence. Funding from the Medical Research Council, British Heart Foundation, Cancer Research UK, Economic and Social Research Council and the NIHR under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Garnett C *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Garnett C, Michie S, West R and Brown J. **Updating the evidence on the effectiveness of the alcohol reduction app, Drink Less: using Bayes factors to analyse trial datasets supplemented with extended recruitment [version 2; peer review: 2 approved]** F1000Research 2019, 8:114 (<https://doi.org/10.12688/f1000research.17952.2>)

First published: 29 Jan 2019, 8:114 (<https://doi.org/10.12688/f1000research.17952.1>)

REVISED Amendments from Version 1

We thank the reviewers for their thorough and thoughtful comments on the paper. We have addressed all these in the revised manuscript and followed the reviewers' suggestions for wording clarifications. In the Results section, we have added a supplementary table of the participant characteristics for those who responded to follow-up (available at [OSF](#)). In the Discussion, we have added a paragraph on the 'value proposition' of the Drink Less app in light of these findings and how the findings informed our decision on which app components to retain or remove. We have also discussed the limitation of the low follow-up rate and an explanation for the difference between the original and extended datasets.

See referee reports

Introduction

A factorial experiment evaluating the effect of 'enhanced' versus 'minimal' versions of five components of the alcohol reduction app, Drink Less, found no clear evidence for simple effects but did find evidence that two-way combinations of certain 'enhanced' components together resulted in greater reductions than 'minimal' versions¹. This was a planned analysis but should be interpreted with caution as the two-way interactive effects were not specifically hypothesised *a priori* and were part of multiple interactions tested. Findings of this sort are not uncommon in experimental studies. One approach is to start another randomised trial specifically to test this hypothesis. A potentially more efficient alternative is to extend the trial with further recruitment and test this and other hypotheses using Bayes factors^{2,3}. We used this approach with the Drink Less app.

Bayes factors are a measure of strength of evidence and allow researchers to 'top-up' their results from one trial with additional data collected, regardless of the stopping rule, unlike frequentist statistics². The use of Bayes factors supports efficient, incremental model building³, as evidence can be continuously accumulated until it is clear whether there is an association or not⁴. The rapid accumulation of large amounts of data about digital behaviour change interventions (DBCI) offers the opportunity to apply emerging methods to their evaluation. DBCIs often have the capacity to continue automatic data collection beyond the end of a trial with little or no additional resources. This paper will illustrate how Bayes factors can be used to optimise a DBCI by updating evidence from an effectiveness trial using the example of Drink Less—an alcohol reduction app.

Bayes factors are the ratio of the average likelihood of two competing hypotheses being correct given a set of data and can overcome some of the issues associated with traditional frequentist statistics⁵. They indicate the relevant strength of evidence for two hypotheses; when evaluating interventions, the two hypotheses are typically the alternative hypothesis (the intervention had the desired effect) and the null hypothesis (the intervention had no effect). Bayes factors, unlike frequentist statistics, can distinguish between two interpretations of a non-significant result: i) support for the null hypothesis of 'no effect' and ii) data are insensitive to detect an effect i.e. 'unsure about the presence of an effect'^{5,6}. Calculating Bayes factors to supplement frequentist

statistics is a quick and simple procedure with several software packages freely available (e.g. an online calculator developed by Zoltan Dienes⁷). Researchers are actively encouraged to supplement, or even replace, classical frequentist hypothesis testing with a Bayesian approach to provide greater interpretative value to any non-significant results⁸. This is important as often non-significant results are misinterpreted as evidence for no effect; a review of trials conducted in addictions research found that the reporting of 'no difference' was only appropriate in a small number of papers reporting this⁹.

The use of Bayes factors also has another major advantage over the traditional frequentist approach that relates to the stopping rule. The traditional frequentist approach necessitates a strict stopping rule and a single analysis of data. Typically, this involves an *a priori* power calculation to specify the required sample size for data collection and the trial to end at that point. Subsequent 'topping-up' of existing data and re-analysing the new larger data set is 'prohibited'¹⁰. This is because any *p*-value between 0 and 1 is equally likely if the null hypothesis is true, regardless of how much data are collected¹¹. Therefore, given enough time and data collection, a significant *p*-value will always be obtained even if the null hypothesis is true¹⁰. So if researchers find a non-significant result—which cannot distinguish between support for the null hypothesis and being insensitive to detect an effect—then a new study would be required to build on these findings. Restarting the process is a waste of research resources but necessary in the context of using a frequentist approach for analysis because additional data collected cannot be analysed. However, this is not the case when using Bayes factors, as they are driven towards zero when the null hypothesis is true and additional data are collected¹⁰. Therefore, researchers may use Bayes factors to analyse additional data to complement an employed stopping rule².

In the evaluation of DBCIs, using Bayes factors is beginning to complement traditional frequentist statistics^{4,12}, and analysing additional data would be of particular benefit. Data collection for a DBCI effectiveness trial is typically automated and therefore does not require additional resources to continue after a pre-specified sample size is reached. Rapid evaluations of DBCIs and efficient accumulation of evidence can be used to inform future versions, keeping pace with advances in technology. Using Bayes factors to update findings about the relative plausibility of the two hypotheses allows researchers to assess the DBCI's effectiveness in an ongoing manner⁴. This remains useful when deciding about whether there is sufficient evidence to demonstrate effectiveness and, therefore, continued development¹³. To the authors' knowledge, no DBCIs have used additional data collected to supplement original effectiveness trial findings and no trials have used Bayes factors to provide further insight based on additional data. However, Bayes factors have been used in trials for superiority, non-inferiority and equivalence designs to allow for explicit quantification of evidence in favour of the null hypothesis¹⁴. Bayesian analyses, more generally, are often used in clinical trials for dose finding, efficacy monitoring, toxicity monitoring, and for diagnosis/decision making¹⁵. For example, Bayesian analyses were used to simultaneously monitor toxicity and efficacy in a parallel phase I/II clinical trial design for combination therapies¹⁶.

DBCI require novel methods of evaluation that are quick and timely to inform the optimisation of the intervention¹⁷. The multiphase optimisation strategy (MOST) is a method for building, optimising and evaluating multicomponent behavioural interventions. It involves a series of steps identifying the set of intervention components to be examined and evaluating the effects of these components^{13,18}. Factorial trial designs allow the simultaneous evaluation of the intervention components, which enables both the independent and interactive effects to be estimated¹³. Using a factorial trial to evaluate a DBCI can overcome some of the challenges associated with using the traditional randomised controlled trial, such as prolonged duration from recruitment to publication and a high-cost trial implementation^{19,20}. The results from a factorial trial can be used to make decisions about which components to retain when optimising the intervention¹⁸.

The Drink Less smartphone app is a DBCI aimed at supporting people who drink excessively to reduce their alcohol consumption. It was developed using evidence and theory, following MOST. The app was analysed in a full factorial trial to assess the effectiveness of its five intervention modules and their effects on app usage and subsequent usability ratings²¹. The stopping rule for data collection, in line with the frequentist approach to analysis, was pre-specified, although data collection continued under the same conditions as the original factorial trial. Analysis of the original trial data using Bayes factors indicated that the data were insensitive to detect main effects but that combinations of the modules appeared effective¹.

Aims

The aims of this study are substantive and methodological:

1. To update the evidence on effectiveness of Drink Less app components singly and in combination. Specifically, what are the main and two-way interactive effects of the intervention modules on:
 - i. Change in weekly alcohol consumption
 - ii. Change in full Alcohol Use Disorders Identification Test (AUDIT) score
2. To demonstrate how Bayes factors can be used to analyse additional outcome data collected in effectiveness trials and update beliefs about hypotheses.

Methods

Design

A between-subject full factorial (2⁵) trial to evaluate the effectiveness of five intervention modules in the Drink Less app. The research questions were specified prior to the trial commencing and pre-registered on ISRCTN (registration number: [ISRCTN40104069](https://www.isrctn.com/ISRCTN40104069)) and published in an open-access protocol paper²¹.

Participants

Participants were included in the study if they: were aged 18 or over; lived in the UK (only available on UK Apple app store

and users had to select 'UK' for 'Country?'); had an AUDIT score of 8 or above (indicative of excessive drinking²²); were interested in reducing their drinking (indicated by the question 'why are you using this app?' with users choosing 'interested in drinking less' over 'just browsing'); provided an email address and had downloaded a 'trial version' of the app (described below).

The sample size for the original factorial trial was 672 providing 80% power (with alpha at 5%, 1:1 allocation and a two-tailed test) to detect a mean change in alcohol consumption of 5 units between the 'enhanced' and 'minimal' versions for each intervention module²³, comparable with a face-to-face brief intervention²⁴. This assumed a mean of 27 weekly units at follow-up in the control group, a mean of 22 units in the intervention group and a SD of 23 units for both (d=0.22).

Recruitment was undertaken via promotion from organisations, such as Public Health England, Cancer Research UK, and listing the app in the iTunes Store according to best practices for app store optimisation.

Measures

Baseline measures included the AUDIT questionnaire and a socio-demographic assessment (age, gender, ethnic group, level of education, employment status and current smoking status). The primary outcome measure was self-reported change in past week alcohol consumption (the difference between one-month follow-up and baseline). Past week alcohol consumption was derived from the frequency (Q1) and quantity (Q2) questions of the AUDIT-Consumption (AUDIT-C) questionnaire. The secondary outcome measure was self-reported change in full AUDIT score; in addition to the three questions on consumption in the AUDIT-C, the full AUDIT includes questions assessing harmful alcohol use (e.g. alcohol-related injuries) and symptoms of dependence. Other secondary outcome measures included in the original, full factorial trial were usage data and usability ratings though were not considered in this paper. Details of these measures are described elsewhere¹, and the data and Bayes Factors calculated are reported on the Open Science Framework (<https://osf.io/kqm8b/>).

Interventions

The Drink Less app is a DBCI for people who drink excessively to help them reduce their alcohol consumption. It is freely available on the UK version of the Apple App Store for all smartphones and tablets running iOS8 or above. The content of the app did not change during the trial except for minor bug fixes (to ensure compatibility with iOS 10).

The app is structured around goal setting: users can set their own goals based on units, cost, alcohol free days or calories with information on the UK drinking guidelines, units and alcohol-related harms. There are five intervention modules that aim to help them achieve their goal: Normative Feedback (providing normative feedback on the user's level of drinking relative to others); Cognitive Bias Re-training (a game to retrain approach-avoidance bias for alcoholic drinks); Self-monitoring and Feedback (providing a facility for self-monitoring of drinking and

receipt of feedback); Action Planning (helping users to undertake action planning to avoid drinking), and Identity Change (promoting a change in identity in relation to alcohol). In the trial version of the app, the five intervention modules existed in two versions: i) an ‘enhanced’ version containing the predicted active ingredients and ii) a ‘minimal’ version that acted as a control.

A detailed description of the content, development and factorial trial evaluation of the app is reported in two separate papers^{1,25}.

Procedures

Data collection for the factorial trial began on 18th May 2016 and the required sample of eligible users was reached on 10th July 2016; follow-up data were collected until 28th August 2016. Trial data was collected continuously for a further four months until 19th December 2016 under the same conditions as the original factorial trial (i.e. a ‘trial version’).

Informed consent to participate in the trial was obtained from all participants on first opening the app. Users who consented to participate completed the AUDIT and a socio-demographic questionnaire, indicated their reason for using the app and provided their email address for follow-up (a prize of £100 was offered in an attempt to decrease the proportion of users leaving this field blank). Users were then provided with their AUDIT score and, those who met the inclusion criteria, were randomised to one of 32 experimental conditions using an automated algorithm within the app for block randomisation.

Follow-up was conducted 28 days after participants downloaded the app and the questionnaire consisted of the full AUDIT and usability measures. Follow-up was conducted in two ways: i) via email with a link to the questionnaire in an online survey tool (Qualtrics), which also sent up to four reminders, and ii) within the app. Participants included according to the original trial and stopping rule were due to complete the follow-up questionnaire up until 29th August 2016 and were contacted via email (through Qualtrics) and the app. Participants due to complete the follow-up questionnaire from 30th August onwards, were only contacted via the app.

Ethical approval

Ethical approval for Drink Less from the UCL Ethics Committee under the ‘optimisation and implementation of interventions to change health-related behaviours’ project (CEHP/2013/508).

Analysis

All analyses were conducted using R version 3.4.0. The analysis plan for this paper followed a similar analysis plan as for the original factorial trial (which was pre-registered on 13th February 2016; ISRCTN40104069²¹).

Participant characteristics were reported descriptively by intervention module. A factorial between-subjects design was used to assess the main and two-way interactive effects of the five intervention modules on the primary and secondary outcome measures. Analyses were conducted amongst responders only, those who completed the follow-up questionnaire. Bayes factors were

calculated for each analysis assessing the main and the two-way interaction effects of the five intervention modules on the outcome measures. The two-way interactions were defined as enhanced/enhanced versus minimal/minimal for each pair of intervention modules. The mean difference and standard error of the mean difference for each main and two-way interactive effect was calculated. A half normal distribution was used to specify the predicted effect. Peak at 0 (no effect) with a SD equal to the expected effect size. This is a conservative approach and represents a hypothesis that the intervention had a least some positive effect, with the effect being more likely to be smaller than larger. Bayes factors were calculated using an online calculator⁷.

The expected effect size for the primary calculation of Bayes factors was a reduction of 5 units per week ($d=0.22$), reflecting a large effect and that of the power calculation for the original factorial trial. Bayes Factors were also calculated for a medium effect (reduction of 3 units per week), and a small effect (reduction of 0.5 units per week) to permit a relative judgment for screening purposes. The expected effect size for the secondary outcome measure was calculated by translating the estimated effect size for the primary outcome measure ($d=0.22$) into the equivalent mean difference score of 1.45 (mean=19.1, SD=6.56 [based on original trial users, $n=672$]). Bayes factors will be interpreted in terms of categories of evidential strength (see Table 1)^{5,26}.

Results

Study sample

The total sample size was 2586, of these 1914 (74.0%) were additional users to the original factorial trial (672, 26.0%). In total, 342 users (13.2%) completed the primary outcome measure in the follow-up questionnaire—the original users’ response rate was 26.6% and the additional users’ response rate was 8.5%. Figure 1 shows a flow chart of users throughout the study.

Socio-demographic and drinking characteristics of participants are reported in Table 2. Participants’ mean age was 37.2 years, 53.4% were women, 95.8% were white, 74.3% had post-16 qualifications, 87.0% were employed, and 30.0% were current

Table 1. Interpretation of Bayes factors.

Bayes factor	Interpretation
>30	Very strong evidence for H_1
10–30	Strong evidence for H_1
3–10	Moderate evidence for H_1
1–3	Anecdotal evidence for H_1
1	No evidence
0.33–1	Anecdotal evidence for H_0
0.10–0.33	Moderate evidence for H_0
0.03–0.10	Strong evidence for H_0
<0.03	Very strong evidence for H_0

H_1 , alternative hypothesis; H_0 , null hypothesis.

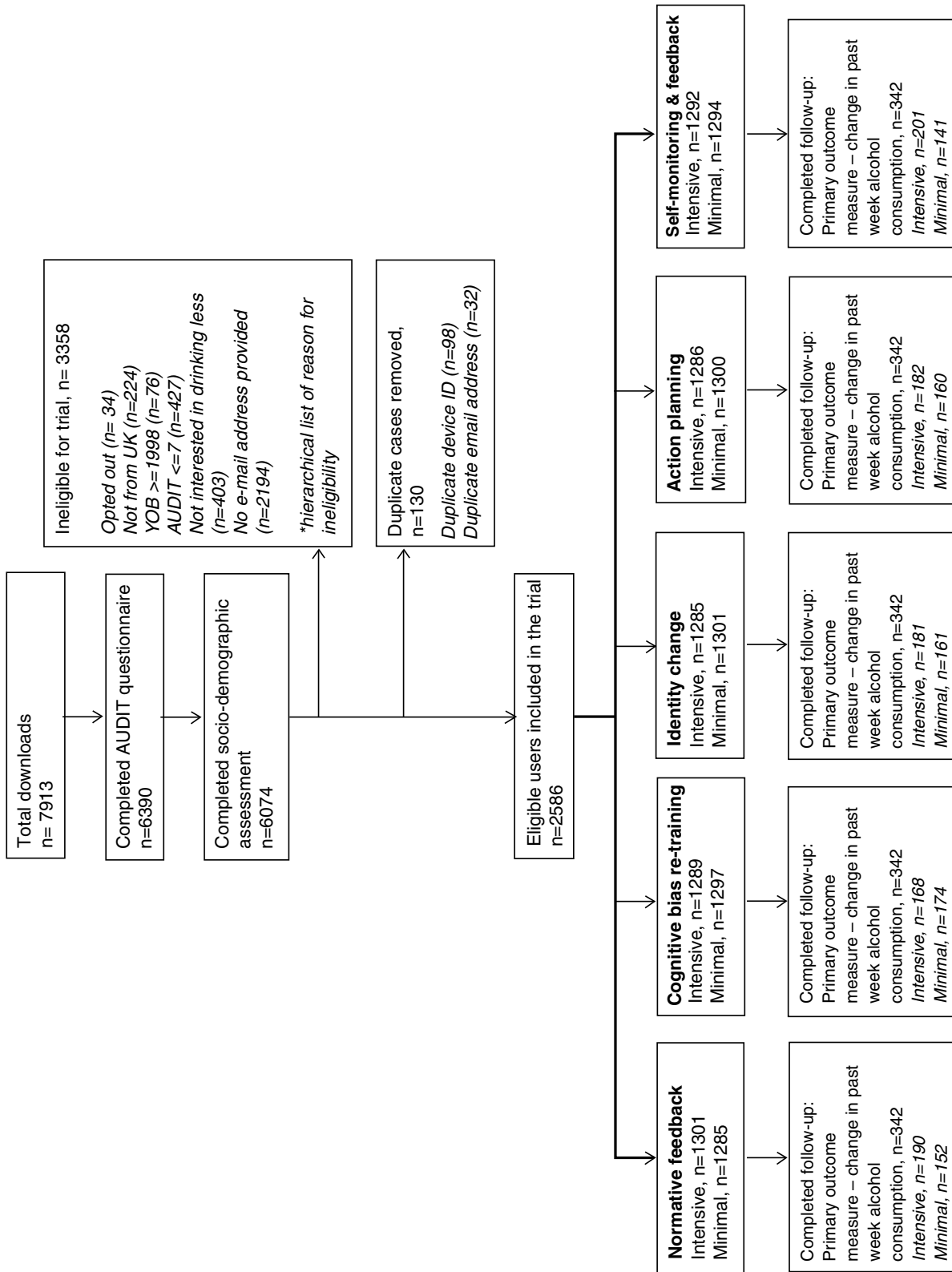


Figure 1. Flow chart of users.

Table 2. Participants' characteristics at baseline. Data given as mean (SD), unless stated.

Variable	All	Normative Feedback		Cognitive Bias Re-training		Self-monitoring & Feedback		Action Planning		Identity Change	
		Enh	Min	Enh	Min	Enh	Min	Enh	Min	Enh	Min
Age	37.2 (10.64)	36.8 (10.49)	37.5 (10.77)	37.0 (10.56)	37.3 (10.72)	37.3 (10.88)	37.0 (10.38)	37.4 (10.87)	37.0 (10.40)	37.4 (10.75)	36.9 (10.52)
% Women (n)	53.4% (1381)	52.7% (686)	54.1% (695)	53.0% (693)	53.8% (688)	53.0% (685)	53.8% (696)	53.5% (688)	53.3% (693)	52.1% (670)	54.7% (711)
% White (n)	95.8% (2477)	96.1% (1250)	95.5% (1227)	96.3% (1241)	95.3% (1236)	95.7% (1237)	95.8% (1240)	95.8% (1232)	95.8% (1245)	95.3% (1224)	96.3% (1253)
% Post-16 qualifications (n)	74.3% (1921)	74.4% (968)	74.2% (953)	74.3% (958)	74.2% (963)	74.1% (958)	74.4% (963)	74.2% (954)	74.4% (967)	74.0% (951)	74.6% (970)
% Employed (n)	87.0% (2250)	87.0% (1132)	87.0% (1118)	87.1% (1123)	86.9% (1127)	85.6% (1106)	88.4% (1144)	85.1% (1095)	88.8% (1155)	86.6% (1113)	87.4% (1137)
% Current smokers (n)	30.0% (776)	29.7% (386)	30.4% (390)	29.8% (384)	30.2% (392)	28.9% (373)	31.1% (403)	29.8% (383)	30.2% (393)	28.0% (360)	32.0% (416)
Past week alcohol consumption (units)	39.0 (26.93)	39.0 (27.02)	39.0 (26.85)	39.7 (27.66)	38.3 (26.18)	39.2 (26.99)	38.8 (26.88)	38.8 (26.32)	39.2 (27.53)	38.7 (26.84)	39.3 (27.03)
AUDIT score	19.1 (6.66)	27.0 (19.18)	26.9 (18.98)	19.4 (6.78)	18.7 (6.52)	19.2 (6.63)	19.0 (6.69)	19.3 (6.67)	18.9 (6.64)	18.9 (6.56)	19.2 (6.76)
AUDIT-C score	9.3 (1.76)	9.3 (1.76)	9.3 (1.77)	9.3 (1.80)	9.3 (1.73)	9.3 (1.78)	9.3 (1.75)	9.3 (1.71)	9.3 (1.82)	9.3 (1.74)	9.3 (1.78)

Enh, enhanced; Min, minimal.

smokers. Mean weekly alcohol consumption was 39.0 units, mean AUDIT-C score was 9.3, and mean AUDIT score was 19.1, indicating harmful drinking. Participants' characteristics by intervention module are reported in [Table 2](#). Generally, characteristics were similar for the enhanced and minimal version of each intervention module. The characteristics of participants who responded to the follow-up questionnaire (n=342) are reported in [Supplementary Table 1](#).

Change in past week's alcohol consumption

The main effects of the intervention modules are reported in [Table 3](#) for the change in past week's alcohol consumption. Bayes factors showed that the data were insensitive to detect an effect for *Normative Feedback* for effect sizes of 5-, 3- and 0.5-unit reductions (0.47<BF<0.97). Data were insensitive to detect an effect for *Cognitive Bias Re-training* for effect sizes of 5-, 3- and 0.5-unit reductions (0.74<BF<1.06). Bayes factors showed that the data were insensitive to detect an effect for *Self-monitoring and Feedback* for effect sizes of 5-, 3- and 0.5-unit reductions (0.43<BF<0.95). Bayes factors showed that the data were insensitive to detect an effect for *Action Planning* for effect sizes of 5-, 3- and 0.5-unit reductions (0.83<BF<1.08). Bayes factors for *Identity Change* showed support for the null hypothesis of no difference between the enhanced and minimal version of the module for a 5-unit reduction (BF=0.22), though data were insensitive to detect an effect for 3- and 0.5-unit reductions (0.34<BF<0.81). The data were insensitive to detect a two-way interactive effect between any pair of intervention modules for effect sizes of 5-, 3- or 0.5-unit reductions (0.35<BF<1.22), except for

between *Self-monitoring and Feedback* and *Identity Change* for a 5-unit reduction which supported the null hypothesis (BF=0.31) (see [Extended data, Supplementary Table 2²⁷](#)).

Change in AUDIT score

The main effects of the intervention modules are reported in [Table 4](#) for the change in AUDIT score. The data were insensitive to detect an effect on change in AUDIT score for: *Normative Feedback* (BF=0.60); *Cognitive Bias Re-training* (BF=0.98); and *Action Planning* (BF=0.95). The data supported evidence for the null hypothesis of no difference in AUDIT score between enhanced and minimal versions of *Self-monitoring and Feedback* (BF=0.15) and *Identity Change* (BF=0.14). The two-way interactive effects of intervention modules on change in AUDIT score (see [Extended data, Supplementary Table 3²⁷](#)) showed that the majority of data were insensitive to detect any two-way interactive effects (0.33<BF<1.99). Data supported the null hypothesis for no difference between enhanced and minimal versions between *Normative Feedback* and *Identity Change* (BF=0.29) and *Self-Monitoring and Feedback* and *Identity Change* (BF=0.18).

Exploratory analysis of a synergistic effect on change in past week's alcohol consumption

Four intervention modules (*Normative Feedback*, *Cognitive Bias Re-Training*, *Self-Monitoring and Feedback*, and *Action Planning*) have some evidence in support of their role of reducing alcohol consumption. Therefore, an additional exploratory analysis was conducted to assess whether there is a larger cumulative effect of the combination of all four modules in the

enhanced version compared with the minimal version. This was done for responders only (n=39; 12 “off” vs 27 “on”) and for last observation carried forward (n=324; 164 “off” vs 160 “on”) to provide potential evidence for what effect size we can expect when planning a definitive trial with longer-term follow-up. Last observation carried forward means that participants’ past week alcohol consumption at follow-up was used for all of those who responded to follow-up and the baseline measure for past week alcohol consumption was used for those who did not respond to follow-up. Whilst last observation carried forward has its limitations, it maintains the variability within the data. Table 5 reports the Bayes factors for these analyses. There was a large numerical difference between all enhanced and all minimal for the four modules amongst responders only, although the Bayes factors found that the data were

insensitive to detect an effect, which may be due in part to the small sample size.

Discussion

The calculation of Bayes factors for additional data collected beyond the original factorial trial of Drink Less has allowed us to accumulate and update existing evidence on the effectiveness of its intervention components in reducing alcohol consumption. The supplemented data remained insensitive to detect whether the Drink Less app components have large (5-unit) individual or two-way interactive effects on reducing alcohol consumption though tended towards anecdotal evidence for the null hypothesis of no effect. There was evidence of two-way interactive effects in the original factorial trial that is no longer supported by the supplemented data.

Table 3. Main effects of intervention modules on change in past week’s alcohol consumption. A negative number indicates a reduction over time.

Variable	Mean change in alcohol consumption, Units (SD)				Bayes factor		
	Enhanced	Minimal	F	P	B _{H(0,5)}	B _{H(0,3)}	B _{H(0,0.5)}
Normative Feedback	-12.5 (25.70)	-12.7 (26.57)	0.007	0.933	0.47	0.66	0.97
Cognitive Bias Re-training	-13.4 (26.93)	-11.9 (25.22)	0.280	0.597	0.74	0.96	1.06
Self-monitoring and Feedback	-12.3 (24.97)	-13.0 (27.61)	0.052	0.820	0.43	0.62	0.95
Action Planning	-13.5 (24.70)	-11.6 (27.55)	0.443	0.506	0.83	1.06	1.08
Identity Change	-10.7 (27.76)	-14.8 (23.89)	2.144	0.144	0.22	0.34	0.81

Table 4. Main effects of intervention modules on change in AUDIT score.

Variable	Mean change in AUDIT score (SD)				Bayes factor
	Enhanced	Minimal	F	P	B _{H(0,1.45)}
Normative Feedback	-2.4 (5.55)	-2.04 (6.11)	0.298	0.586	0.60
Cognitive Bias Re-training	-2.5 (5.73)	-1.9 (5.88)	1.042	0.308	0.98
Self-monitoring and Feedback	-1.8 (5.62)	-2.8 (6.02)	3.006	0.084	0.15
Action Planning	-2.5 (6.06)	-1.9 (5.50)	0.983	0.322	0.95
Identity Change	-1.7 (5.89)	-2.8 (5.66)	3.529	0.061	0.14

Table 5. Four modules in ‘enhanced’ vs four modules in ‘minimal’ versions for past week alcohol consumption.

Variable	PWAC, Units (SD)				Bayes factor		
	All enhanced	All minimal	F	P	B _{H(0,5)}	B _{H(0,3)}	B _{H(0,0.5)}
Responders only (change in PWAC)	-21.6 (20.36)	-12.1 (26.82)	1.474	0.232	1.42	1.29	1.05
Last observation carried forward (PWAC)	36.7 (28.48)	37.4 (26.59)	0.059	0.808	0.62	0.82	1.02

PWAC, past week alcohol consumption.

The current data also remained insensitive to detect whether the four most promising components (*Normative Feedback*, *Cognitive Bias Re-Training*, *Self-Monitoring and Feedback* and *Action Planning*) may each have effects smaller than 5 units. An unplanned analysis provided weak anecdotal evidence of a synergistic effect of the ‘enhanced’ versions of these four intervention modules together. On both past week alcohol consumption and AUDIT score, and across several alternative effect sizes, there was support for no effect of the fifth intervention module, *Identity Change*. These findings, alongside results from analysing user feedback and usage data on the most frequently visited screens, guided the decision to remove the *Identity Change* module from the next major app update whilst retaining *Normative Feedback* and *Cognitive Bias Re-Training*, and *Self-Monitoring and Feedback* and *Action Planning*.

Whilst this study did not find evidence of a large individual effect of any of the intervention modules, there remains some evidence to suggest that an optimised version of the app (with the removal of the *Identity Change* module) may yet prove effective. As with the original factorial trial, there are concerns that the minimal versions were too active in an attempt to promote engagement amongst all participants. Even participants who were randomised to receive the minimal versions of every intervention module were able to set goals and track their drinks, which is associated with reduced consumption²⁸. Most alcohol reduction apps include few techniques to change behaviour²⁹ suggesting that even the minimal version of Drink Less was more active than most existing alcohol reduction apps. Therefore, effectiveness estimates derived from this approach are likely to be conservative. Furthermore, Drink Less users have excellent levels of engagement with the app³⁰, which is necessary (but not sufficient) for an intervention to be effective. Additionally, a content analysis of user feedback (available as a short report here: <https://osf.io/d3w8tr/>) found that of the ‘Information giving’ category, the majority provided positive feedback on the app as a whole. A sample of the user feedback is available to view on the Drink Less website³¹. Drink Less is also one of the leading alcohol reduction apps in the UK with over 50,000 unique users and an average 4.1-star rating (as of June 2019).

A major strength of this study is its illustration of how it is possible to evaluate data from trials of DBCIs in an on-going manner. No additional resources were required to continue data collection within the original trial of Drink Less as the app remained freely available on the UK Apple app store and the notification to complete the follow-up questionnaire had already been programmed. Analysing the supplemented dataset has allowed us to update our findings and provided more confidence in our original decisions on which components to retain or remove as part of the process of optimising the intervention¹⁸ to improve its effectiveness and usability. We are also much clearer that any definitive trial must be powered to detect small effects and designed to inform a pragmatic decision about whether to invest resources in recommending the app. The optimisation of the Drink Less intervention was based on the findings from this study as well as on user feedback and findings from a meta-analysis of the intervention components in digital alcohol interventions associated

with effectiveness³². The findings from this study informed the removal of the ‘Identity Change’ module and retention of the remaining four modules.” The stopping rule in frequentist statistics means that additional trial data collected as part of an effectiveness trial for a DBCI would go to waste. The use of Bayes factors in this situation prevents unnecessary waste of resources and enables researchers to continually update their evidence on a DBCI rather than collect and analyse individual data sets as part of separate trials.

A limitation of this study and the use of Bayes factors was that we were not able to use the intention-to-treat (ITT) approach in the analysis (as was done for the original trial), whereby those lost to follow-up (non-responders) were assumed to be drinking at baseline levels. Whilst Bayes factors can overcome a lot of the issues with the frequentist approach, they are not meaningful when assumptions are made that limit the variability in the data. Due to low overall follow-up rates (13.2%) in this larger sample, the ITT assumption that there was no change in the large majority of the sample drives the variability down, which in turn drives support for the null hypothesis. This highlights that Bayes factors were not useful in this study when using the ITT assumption, which limits the variability in the data.

We acknowledge that the follow-up rate is very low and this is likely to be due to the lack of financial incentive for completing the follow-up survey, which are known to increase response rates in randomised trials³³. Furthermore, the follow-up rate in the extended dataset was lower than for the original trial dataset; this is likely because participants were only contacted via the app for the extended dataset whilst the participants in the original dataset were also contacted via email.

The intervention modules of the Drink Less app do not have a large individual effect on reducing alcohol-related outcomes, though they may have a small effect that the current data were unable to detect. There is weak evidence for a synergistic effect of the ‘enhanced’ versions of four intervention modules together: *Normative Feedback* and *Cognitive Bias Re-Training*, and *Self-Monitoring and Feedback* and *Action Planning*. This study has updated the existing evidence on the effectiveness of intervention modules in the Drink Less app. In the event of uncertain results following a primary analysis, Bayes factors can be used to ‘top-up’ results from DBCI trials with any additional data collected, therefore supporting efficient, incremental model building to inform decision-making.

Data availability

Underlying data

A dataset containing the extended trial outcomes is available on OSF. DOI: <https://doi.org/10.17605/OSF.IO/KQM8B27>.

Extended data

Extended data are available on OSF. DOI: <https://doi.org/10.17605/OSF.IO/KQM8B27>.

Supplementary Table 1. Participants’ characteristics (who responded to follow-up, n=342) at baseline.

Supplementary Table 2. Two-way interactive effects of intervention modules on change in past week's alcohol consumption.

Supplementary Table 3. Two-way interactive effects of intervention modules on change in AUDIT score.

Data are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information

JB and RW are funded by Cancer Research UK (CRUK; C1417/A22962). CG and SM are funded by CRUK and the NIHR School for Public Health Research (SPHR). Drink Less was funded by the NIHR SPHR, the UK Centre for Tobacco

and Alcohol Studies (UKCTAS), the Society for the Study of Addiction (SSA), and CRUK. The views expressed are those of the authors and not necessarily those of the the NIHR or the Department of Health and Social Care.

The research team is part of the UKCTAS, a UKCRC Public Health Research Centre of Excellence. Funding from the Medical Research Council, British Heart Foundation, Cancer Research UK, Economic and Social Research Council and the NIHR under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

The funders played no role in the design, conduct or analysis of the study, nor in the interpretation or reporting of study findings.

References

- Crane D, Garnett C, Michie S, *et al.*: **A smartphone app to reduce excessive alcohol consumption: Identifying the effectiveness of intervention components in a factorial randomised control trial.** *Sci Rep.* 2018; **8**(1): 4384. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rouder JN: **Optional stopping: No problem for Bayesians.** *Psychon Bull Rev.* 2014; **21**(2): 301–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Patrick K, Hekler EB, Estrin D, *et al.*: **The Pace of Technologic Change: Implications for Digital Health Behavior Intervention Research** *Am J Prev Med.* 2016; **51**(5): 816–24. [PubMed Abstract](#) | [Publisher Full Text](#)
- West R, Michie S: **A Guide to Development and Evaluation of Digital Interventions in Healthcare.** London: Silverback Publishing; 2016.
- Jeffreys H: **The Theory of Probability.** 3rd ed. Oxford University Press; 1961. [Reference Source](#)
- Dienes Z: **Using Bayes to get the most out of non-significant results.** *Front Psychol.* 2014; **5**: 781. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dienes Z: **Making the most of your data with Bayes.** [Internet]. [Reference Source](#)
- Berger J: **The case for objective Bayesian analysis.** *Bayesian Anal.* 2006; **1**(3): 385–402. [Reference Source](#)
- Beard E, Dienes Z, Muirhead C, *et al.*: **Using Bayes factors for testing hypotheses about intervention effectiveness in addictions research.** *Addiction.* 2016; **111**(12): 2230–47. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dienes Z: **Bayesian Versus Orthodox Statistics: Which Side Are You On?** *Perspect Psychol Sci.* 2011; **6**(3): 274–90. [PubMed Abstract](#) | [Publisher Full Text](#)
- Rouder JN, Speckman PL, Sun D, *et al.*: **Bayesian t tests for accepting and rejecting the null hypothesis.** *Psychon Bull Rev.* 2009; **16**(2): 225–37. [PubMed Abstract](#) | [Publisher Full Text](#)
- Heino MTJ, Vuorinen M, Hankonen N: **Bayesian evaluation of behavior change interventions: a brief introduction and a practical example.** *Heal Psychol Behav Med.* 2018; **6**(1): 49–78. [Publisher Full Text](#)
- Collins LM, Murphy SA, Strehler V: **The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions.** *Am J Prev Med.* 2007; **32**(5 Suppl): S112–8. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- van Ravenzwaaij D, Menden R, Tendeiro JN, *et al.*: **Bayes factors for superiority, non-inferiority, and equivalence designs.** *BMC Med Res Methodol.* 2019; **19**(1): 71. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lee JJ, Chu CT: **Bayesian clinical trials in action.** *Stat Med.* 2012; **31**(25): 2955–72. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Huang X, Biswas S, Oki Y, *et al.*: **A parallel phase I/II clinical trial design for combination therapies.** *Biometrics.* 2007; **63**(2): 429–36. [PubMed Abstract](#) | [Publisher Full Text](#)
- Collins LM, Trail JB, Kugler KC, *et al.*: **Evaluating individual intervention components: making decisions based on the results of a factorial screening experiment.** *Transl Behav Med.* 2014; **4**(3): 238–51. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Collins LM, Baker TB, Mermelstein RJ, *et al.*: **The multiphase optimization strategy for engineering effective tobacco use interventions.** *Ann Behav Med.* 2011; **41**(2): 208–26. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Murray E, Hekler EB, Andersson G, *et al.*: **Evaluating Digital Health Interventions: Key Questions and Approaches.** *Am J Prev Med.* 2016; **51**(5): 843–51. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pham Q, Wiljer D, Cafazzo JA: **Beyond the Randomized Controlled Trial: A Review of Alternatives in mHealth Clinical Trial Methods.** *JMIR mHealth uHealth.* 2016; **4**(3): e107. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Garnett C, Crane D, Michie S, *et al.*: **Evaluating the effectiveness of a smartphone app to reduce excessive alcohol consumption: protocol for a factorial randomised control trial.** *BMC Public Health.* 2016; **16**(1): 536. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Reinert DF, Allen JP: **The alcohol use disorders identification test: an update of research findings.** *Alcohol Clin Exp Res.* 2007; **31**(2): 185–99. [PubMed Abstract](#) | [Publisher Full Text](#)
- Kunz FM Jr, French MT, Bazargan-Hejazi S: **Cost-effectiveness analysis of a brief intervention delivered to problem drinkers presenting at an inner-city hospital emergency department.** *J Stud Alcohol.* 2004; **65**(3): 363–70. [PubMed Abstract](#) | [Publisher Full Text](#)
- Kaner E, Beyer F, Dickinson HO, *et al.*: **Effectiveness of brief alcohol interventions in primary care populations.** *Cochrane Database Syst Rev.* 2007; **(2)**: CD004148. [PubMed Abstract](#) | [Publisher Full Text](#)
- Garnett C, Crane D, West R, *et al.*: **The development of Drink Less: an alcohol reduction smartphone app for excessive drinkers.** *Transl Behav Med.* 2019; **9**(2): 296–307. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Andraszewicz S, Scheibehenne B, Rieskamp J, *et al.*: **An Introduction to Bayesian Hypothesis Testing for Management Research.** *J Manage.* 2015; **41**(2): 521–43. [Publisher Full Text](#)
- Garnett C, Michie S, West R: **Updating evidence of effectiveness of Drink Less using Bayes factors to analyse additional outcome data.** 2019. <http://www.doi.org/10.17605/OSF.IO/KQM8B>
- Helzer JE, Badger GJ, Rose GL, *et al.*: **Decline in alcohol consumption during two years of daily reporting.** *J Stud Alcohol.* Rutgers University Piscataway, NJ; 2002; **63**(5): 551–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Crane D, Garnett C, Brown J, *et al.*: **Behavior change techniques in popular**

- alcohol reduction apps: content analysis.** *J Med Internet Res.* 2015; 17(5): e118.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Garnett C, Perski O, Tombor I, *et al.*: **Predictors of Engagement, Response to Follow Up, and Extent of Alcohol Reduction in Users of a Smartphone App (Drink Less): Secondary Analysis of a Factorial Randomized Controlled Trial.** *JMIR Mhealth Uhealth.* 2018; 6(12): e11175.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. **Drink Less Testimonials [Internet].**
[Reference Source](#)
32. Garnett CV, Crane D, Brown J, *et al.*: **Behavior Change Techniques Used in Digital Behavior Change Interventions to Reduce Excessive Alcohol Consumption: A Meta-regression.** *Ann Behav Med.* 2018; 52(6): 530–43.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Brueton VC, Tierney JF, Stenning S, *et al.*: **Strategies to improve retention in randomised trials: a Cochrane systematic review and meta-analysis.** *BMJ Open.* 2014; 4(2): e003821.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 16 July 2019

<https://doi.org/10.5256/f1000research.21823.r51200>

© 2019 Heather N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Nick Heather 

Department of Psychology, Faculty of Health & Life Sciences, Northumbria University, Newcastle upon Tyne, UK

The authors have responded satisfactorily to criticisms of the previous version of this paper. I have no further comments to make.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I have long experience in the area of research on brief interventions for hazardous and harmful alcohol consumption. However, although I have authored a publication using the Bayesian approach to hypothesis testing, I am by no means an expert on the use of Bayesian statistics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 16 July 2019

<https://doi.org/10.5256/f1000research.21823.r51199>

© 2019 Leightley D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Daniel Leightley 

King's Centre for Military Health Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

The authors have suitably addressed my concerns. I have no outstanding issues that need to be addressed.

Competing Interests: Researcher on the InDEx app project - an app designed to help armed forces personnel monitor their alcohol consumption

Reviewer Expertise: Design and development of mobile health solutions for mental health.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 13 March 2019

<https://doi.org/10.5256/f1000research.19633.r44925>

© 2019 Heather N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Nick Heather 

Department of Psychology, Faculty of Health & Life Sciences, Northumbria University, Newcastle upon Tyne, UK

This paper reports the results of a factorial experiment evaluating the Drink Less Smart phone app and its behaviour change components, with the benefit of an extended data-set. The aims of the study were both substantive and methodological. In relation to the latter, the paper provides a very useful example of the advantages of Bayesian hypothesis testing and in particular its legitimate provision for extending data collection until a firm conclusion has been reached. The paper also provides a good example, with earlier papers, of the MOST method for developing and evaluating behaviour change interventions.

Regarding the substantive aim, I have two major and several minor comments.

MAJOR:

1. Analysis of the full data-set after extension confirmed the mainly inconclusive results of tests of the 5 individual components of the app in the original evaluation study and failed to confirm the previously reported evidence in favour of two interaction effects between components. In view of the authors' thorough and painstaking development of the app over the years, their rigorous evaluations of the components of the app and their stated intention to optimise the app for a definitive test in an RCT with extended follow-up, these results must be regarded as very disappointing. Yet this is not commented on and no possible explanation is offered why components whose inclusion was supported strongly by theory and previous research failed to show effects on drinking behaviour. Is it possible, as the authors have previously suggested, that the 'minimal', control interventions were too active to allow an effect to emerge? What other explanations are there for these disappointing results? Overall, where do the authors go from here in the attempt to bring this very promising intervention technology to practical use?
2. In the Abstract the authors state: 'There was weak evidence for a synergistic effect of four components'. I feel even this is too strong. In the text on p.9 we have: 'An unplanned analysis

provided weak anecdotal evidence of a synergistic effect of the 'enhanced' versions of these four intervention modules together'. Something along these lines, with the inclusion of 'unplanned' and 'anecdotal' would be more appropriate for the Abstract.

Another concern here is that the putative effect in question derives from a post hoc hypothesis based on unplanned comparisons arrived at only after the extension of data collection. In a frequentist approach to hypothesis testing, this would not of course be permissible but, even in the Bayesian approach, there must surely be some constraints on the legitimacy of testing post hoc hypotheses derived from exploratory analyses (rather than using such analyses to generate hypotheses not tested in the present data). The authors should comments on this issue and, if necessary, seek expert advice.

MINOR:

1. Why was the follow-up rate so much lower in the extended data-set than in the original data (8.5% versus 26.6%)? Can the authors attempt to explain this difference?
2. p.8: '... to provide potential evidence for what effect size we can expect when planning the trial.' This is presumably the definitive trial with extended follow-up but this is not clear.
3. What about the data on secondary outcomes? These are not reported here but it is not stated that they will not be considered in this paper and the reader is not told where they will be found.
4. The primary outcome measure of self-reported change in past week alcohol consumption was presumably based on the AUDIT-C questionnaire, as suggested by Table 2, but this is not made clear in the text.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I have long experience in the area of research on brief interventions for hazardous and harmful alcohol consumption. However, although I have authored a publication using the Bayesian approach to hypothesis testing, I am by no means an expert on the use of Bayesian statistics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 04 Jul 2019

Claire Garnett, University College London, London, UK

Analysis of the full data-set after extension confirmed the mainly inconclusive results of tests of the 5 individual components of the app in the original evaluation study and failed to confirm the previously reported evidence in favour of two interaction effects between components. In view of the authors' thorough and painstaking development of the app over the years, their rigorous evaluations of the components of the app and their stated intention to optimise the app for a definitive test in an RCT with extended follow-up, these results must be regarded as very disappointing. Yet this is not commented on and no possible explanation is offered why components whose inclusion was supported strongly by theory and previous research failed to show effects on drinking behaviour. Is it possible, as the authors have previously suggested, that the 'minimal', control interventions were too active to allow an effect to emerge? What other explanations are there for these disappointing results? Overall, where do the authors go from here in the attempt to bring this very promising intervention technology to practical use?

This is a very good point made by both reviewers that we do not sufficiently discuss the value proposition of Drink Less in our discussion section. We have added a paragraph on this:

"Whilst this study did not find evidence of a large individual effect of any of the intervention modules, there remains some evidence to suggest that an optimised version of the app (with the removal of the 'Identity Change' module) may yet prove effective. As with the original factorial trial, there are concerns that the minimal versions were too active in an attempt to promote engagement amongst all participants. Even participants who were randomised to receive the minimal versions of every intervention module were able to set goals and track their drinks, which is associated with reduced consumption [1]. Most alcohol reduction apps include few techniques to change behaviour [2] suggesting that even the minimal version of Drink Less was more active than most existing alcohol reduction apps. Therefore, effectiveness estimates derived from this approach are likely to be conservative. Furthermore, Drink Less users have excellent levels of engagement with the app [3], which is necessary (but not sufficient) for an intervention to be effective. Additionally, a content analysis of user feedback (available as a short report here: <https://osf.io/d3w8r/>) found that of the 'Information giving' category, the majority provided positive feedback on the app as a whole. A sample of the user feedback is available to view on the Drink Less website [4]. Drink Less is also one of the leading alcohol reduction apps in the UK with over 50,000 unique users to date with an average 4.1-star rating (as of June 2019)."

In the Abstract the authors state: 'There was weak evidence for a synergistic effect of four components'. I feel even this is too strong. In the text on p.9 we have: 'An unplanned analysis provided weak anecdotal evidence of a synergistic effect of the 'enhanced' versions of these four intervention modules together'. Something along these lines, with the inclusion of 'unplanned' and 'anecdotal' would be more appropriate for the Abstract.

We have re-worded the abstract to be cautious with our conclusions:

"In an additional exploratory analysis, participants receiving four of the components averaged a numerically greater reduction in consumption than those not receiving any (21.6 versus 12.1 units), but the data were insensitive (BF=1.42).

Data from extended recruitment in a factorial experiment evaluating components of the Drink Less app remained insensitive but tended towards individual and pairs of components not having a large

effect. In an exploratory analysis, there was weak anecdotal evidence for a synergistic effect of four components. In the event of uncertain results, calculating BFs can be used to update the strength of evidence of a dataset supplemented with extended recruitment.”

Another concern here is that the putative effect in question derives from a post hoc hypothesis based on unplanned comparisons arrived at only after the extension of data collection. In a frequentist approach to hypothesis testing, this would not of course be permissible but, even in the Bayesian approach, there must surely be some constraints on the legitimacy of testing post hoc hypotheses derived from exploratory analyses (rather than using such analyses to generate hypotheses not tested in the present data). The authors should comment on this issue and, if necessary, seek expert advice.

These analyses were exploratory rather than testing a post-hoc hypothesis and the Bayesian approach does not have the same limitations as the frequentist approach in conducting exploratory analyses. Also, the conclusions match the strength of the evidence for this exploratory analysis (weak and anecdotal). We plan to pre-register any hypotheses and analysis plans in future studies. Furthermore, a key part of this study was for it to be informative in making decisions on which components to retain in an optimised version of the app and in terms of conducting a power analysis for a definitive trial.

Why was the follow-up rate so much lower in the extended data-set than in the original data (8.5% versus 26.6%)? Can the authors attempt to explain this difference?

We have included a discussion of this in the limitations:

“We acknowledge that the follow-up rate is very low and this is likely to be due to the lack of financial incentive for completing the follow-up survey, which are known to increase response rates in randomised trials (Bructon et al., 2014). Furthermore, the follow-up rate in the extended dataset was lower than for the original trial dataset; this is likely because participants were only contacted via the app for the extended dataset whilst the participants in the original dataset were also contacted via email.”

p.8: ‘... to provide potential evidence for what effect size we can expect when planning the trial.’ This is presumably the definitive trial with extended follow-up but this is not clear.

We have clarified this:

“...to provide potential evidence for what effect size we can expect when planning a definitive trial with longer-term follow-up.”

What about the data on secondary outcomes? These are not reported here but it is not stated that they will not be considered in this paper and the reader is not told where they will be found.

The data on the other secondary outcomes (usage data and usability ratings) were not reported in this paper though we have added information on where they will be found:

“Other secondary outcome measures included in the original, full factorial trial were usage data and usability ratings though were not considered in this paper. Details of these measures are described elsewhere 1, and the data and Bayes Factors calculated are reported on the Open Science Framework (<https://osf.io/kqm8b/>).”

The primary outcome measure of self-reported change in past week alcohol consumption was presumably based on the AUDIT-C questionnaire, as suggested by Table 2, but this is not made clear in the text.

We have clarified this in the measures section:

“The primary outcome measure was self-reported change in past week alcohol consumption (the

difference between one-month follow-up and baseline). Past week alcohol consumption was derived from the frequency (Q1) and quantity (Q2) questions of the AUDIT-Consumption (AUDIT-C) questionnaire.”

Competing Interests: No competing interests were disclosed.

Reviewer Report 11 February 2019

<https://doi.org/10.5256/f1000research.19633.r43750>

© 2019 Leightley D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Daniel Leightley

King's Centre for Military Health Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

This paper describes the effectiveness of DrinkLess, a mobile alcohol intervention, using Bayes factor to further compliment previously published findings. I have one major comment and several minor comments.

Major Comment:

1. The work extends our understanding of DrinkLess and its effectiveness in managing alcohol misuse; however, it would be helpful to make a clear statement on how the Bayes analysis has improved the value proposition of DrinkLess.

Minor Comments:

1. Abstract: “Amongst responders only” – is that the sample who took part in the follow-up questionnaire?;
2. Abstract: “Unplanned comparison” appears to convey a negative connotation the authors could alter to “additional analyses”;
3. Abstract: “four most promising” could be misleading as you only had five components but we also have to be mindful that the data was insensitive;
4. Abstract: “reminded insensitive but tended” are you able to provide any BF for this statement?;
5. Introduction: It would be helpful to provide some discussion (specific examples) on how Bayes have been used in other domains to provide more insight by means of additional data;
6. Methods – Participants: “were interested in reducing their drinking” how was this measured? Were participants research aware, or were they targeted because they had previously stated an interest in reducing alcohol? Or could it be by downloading DrinkLess they were assumed to be interested in reducing their alcohol consumption?;
7. Methods – Participants: Was a geolocation restriction placed on participants? How can you be sure that users were from the UK?;
8. Methods – Intervention: What were the minor bug fixes, is a summary able to be provided as a supplement?;
9. Results: Results present AUDIT-C score, however this is not discussed previously.
10. Results: It would be helpful to have Table 2 represented as supplementary material for those who took part in follow-up;

11. Results: It would be of interest to discuss further the difference in AUDIT and AUDIT-C score and the role the final two questions (risk taking etc) play;
12. Discussion: “no additional resources were required” – is this the case, was the app provisioned for longer than anticipated?;
13. Discussion: “our decision on which components to retain or remove” – a bit more discussion round this aspect would be helpful to the reader;
14. Discussion: A 13.2% follow-up rate appears to be very low, do the authors have any reasons for this?;

Is the rationale for developing the new method (or application) clearly explained?

Partly

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: Researcher on the InDEX app project - an app designed to help armed forces personnel monitor their alcohol consumption

Reviewer Expertise: Mobile health with a focus on alcohol misuse

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 04 Jul 2019

Claire Garnett, University College London, London, UK

The work extends our understanding of DrinkLess and its effectiveness in managing alcohol misuse; however, it would be helpful to make a clear statement on how the Bayes analysis has improved the value proposition of DrinkLess.

This is a very good point made by both reviewers that we do not sufficiently discuss the value proposition of Drink Less in our discussion section. We have added a paragraph on this: “Whilst this study did not find evidence of a large individual effect of any of the intervention modules, there remains some evidence to suggest that an optimised version of the app (with the removal of the ‘Identity Change’ module) may yet prove effective. As with the original factorial trial,

there are concerns that the minimal versions were too active in an attempt to promote engagement amongst all participants. Even participants who were randomised to receive the minimal versions of every intervention module were able to set goals and track their drinks, which is associated with reduced consumption [1]. Most alcohol reduction apps include few techniques to change behaviour [2] suggesting that even the minimal version of Drink Less was more active than most existing alcohol reduction apps. Therefore, effectiveness estimates derived from this approach are likely to be conservative. Furthermore, Drink Less users have excellent levels of engagement with the app [3], which is necessary (but not sufficient) for an intervention to be effective. Additionally, a content analysis of user feedback (available as a short report here: <https://osf.io/d3w8r/>) found that of the 'Information giving' category, the majority provided positive feedback on the app as a whole. A sample of the user feedback is available to view on the Drink Less website [4]. Drink Less is also one of the leading alcohol reduction apps in the UK with over 50,000 unique users to date with an average 4.1-star rating (as of June 2019)."

Abstract: "Amongst responders only" – is that the sample who took part in the follow-up questionnaire?;

Yes, this has been clarified in the abstract with "...amongst responders only (those who completed the questionnaire)."

Abstract: "Unplanned comparison" appears to convey a negative connotation the authors could alter to "additional analyses";

We have reworded this to "additional exploratory analysis" throughout the manuscript.

Abstract: "four most promising" could be misleading as you only had five components but we also have to be mindful that the data was insensitive;

We have re-worded the abstract to be cautious with our conclusions:

"In an additional exploratory analysis, participants receiving four of the components averaged a numerically greater reduction in consumption than those not receiving any (21.6 versus 12.1 units), but the data were insensitive (BF=1.42).

Data from extended recruitment in a factorial experiment evaluating components of the Drink Less app remained insensitive but tended towards individual and pairs of components not having a large effect. In an exploratory analysis, there was weak anecdotal evidence for a synergistic effect of four components. In the event of uncertain results, calculating BFs can be used to update the strength of evidence of a dataset supplemented with extended recruitment."

Abstract: "reminded insensitive but tended" are you able to provide any BF for this statement?;

The relevant BFs for this conclusion are included in the results section of the abstract and we have added the BF range for the two-way interactive effects: "Data were mainly insensitive but tended to support there being no large main effects of the enhanced version of individual components on consumption (0.22

Introduction: It would be helpful to provide some discussion (specific examples) on how Bayes have been used in other domains to provide more insight by means of additional data;

We have added further details into the introduction about previous uses of Bayes factors and Bayesian analyses:

"To the authors' knowledge, no DBCIs have used additional data collected to supplement original effectiveness trial findings and no trials have used Bayes Factors to provide further insight based on additional data. However, Bayes Factors have been used in trials for superiority, non-inferiority and equivalence designs to allow for explicit quantification of evidence in favour of the null

hypothesis [1]. Bayesian analyses, more generally, are often used in clinical trials for dose finding, efficacy monitoring, toxicity monitoring, and for diagnosis/decision making [2]. For example, Bayesian analyses were used to simultaneously monitor toxicity and efficacy in a parallel phase I/II clinical trial design for combination therapies [3].”

Methods – Participants: “were interested in reducing their drinking” how was this measured? Were participants research aware, or were they targeted because they had previously stated an interest in reducing alcohol? Or could it be by downloading DrinkLess they were assumed to be interested in reducing their alcohol consumption?;

Methods – Participants: Was a geolocation restriction placed on participants? How can you be sure that users were from the UK?;

We have clarified details on the participants in the methods section:

“Participants were included in the study if they: were aged 18 or over; lived in the UK (only available on UK Apple app store and users had to select ‘UK’ for ‘Country?’); had an AUDIT score of 8 or above (indicative of excessive drinking); were interested in reducing their drinking (indicated by the question ‘why are you using this app?’ with users choosing ‘interested in drinking less’ over ‘just browsing’); provided an email address and had downloaded a ‘trial version’ of the app (described below).”

Methods – Intervention: What were the minor bug fixes, is a summary able to be provided as a supplement?;

As the bug fixes are minor, we have included a brief explanation in the manuscript: “The content of the app did not change during the trial except for minor bug fixes (to ensure compatibility with iOS 10).”

Results: Results present AUDIT-C score, however this is not discussed previously.

We have clarified this in the measures section:

“The primary outcome measure was self-reported change in past week alcohol consumption (the difference between one-month follow-up and baseline). Past week alcohol consumption was derived from the frequency (Q1) and quantity (Q2) questions of the AUDIT-Consumption (AUDIT-C) questionnaire.”

Results: It would be helpful to have Table 2 represented as supplementary material for those who took part in follow-up;

We have added a supplementary table of the participant characteristics for those who responded to follow-up.

Results: It would be of interest to discuss further the difference in AUDIT and AUDIT-C score and the role the final two questions (risk taking etc) play;

We have discussed the differences between the AUDIT-C and AUDIT scores in the methods section:

“The secondary outcome measure was self-reported change in full AUDIT score; in addition to the three questions on consumption in the AUDIT-C, the full AUDIT includes questions assessing harmful alcohol use (e.g. alcohol-related injuries) and symptoms of dependence.”

Discussion: “no additional resources were required” – is this the case, was the app provisioned for longer than anticipated?;

The app was always intended to be available for the long-term (i.e. not removing it after completion of the trial). We have added more explanation to this section of the discussion:

“No additional resources were required to continue data collection within the original trial of Drink Less as the app remained freely available on the UK Apple app store and the notification to complete the follow-up questionnaire had already been programmed.”

Discussion: “our decision on which components to retain or remove” – a bit more discussion round this aspect would be helpful to the reader;

We have elaborated on this in the discussion:

“Analysing the supplemented dataset has allowed us to update our findings and provided more confidence in our original decisions on which components to retain or remove as part of the process of optimising the intervention ¹⁵ to improve its effectiveness and usability. We are also much clearer that any definitive trial must be powered to detect small effects and designed to inform a pragmatic decision about whether to invest resources in recommending the app. The optimisation of the Drink Less intervention was based on the findings from this study as well as on user feedback and findings from a meta-analysis of the intervention components in digital alcohol interventions associated with effectiveness [9]. The findings from this study informed the removal of the ‘Identity Change’ module and retention of the remaining four modules.”

Discussion: A 13.2% follow-up rate appears to be very low, do the authors have any reasons for this?;

We have included a discussion of this in the limitations:

“We acknowledge that the follow-up rate is very low and this is likely to be due to the lack of financial incentive for completing the follow-up survey, which are known to increase response rates in randomised trials (Brueton et al., 2014). Furthermore, the follow-up rate in the extended dataset was lower than for the original trial dataset; this is likely because participants were only contacted via the app for the extended dataset whilst the participants in the original dataset were also contacted via email.”

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research