ADDICTION

RESEARCH REPORT

doi:10.1111/add.15327

Effect of alcohol label designs with different pictorial representations of alcohol content and health warnings on knowledge and understanding of low-risk drinking guidelines: a randomized controlled trial

Natalie Gold[1,2], Mark Egan[3], Kristina Londakova[3], Abigail Mottershaw[3], Hugo Harper[3], Robyn Burton[1,4], Clive Henn[1], Maria Smolar[1], Matthew Walmsley[1], Rohan Arambepola[1,5,6], Robin Watson[1,7], Sarah Bowen[1,8] & Felix Greaves[1,9]

Public Health England, London, UK,[1] Department of Philosophy, University of Oxford, Oxford, UK,[2] Behavioural Insights Team, London, UK,[3] Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK,[4] Oxford Big Data Institute, Oxford, UK,[5] Nuffield Department of Medicine, University of Oxford, Oxford, UK,[6] Department of Anthropology, Durham University, Durham, UK,[7] School of Economics Nottingham, UK[8] and Department of Primary Care and Public Health, Imperial College, London, UK[9]

ABSTRACT

Background and aims The UK low-risk drinking guidelines (LRDG) recommend not regularly drinking more than 14 units of alcohol per week. We tested the effect of different pictorial representations of alcohol content, some with a health warning, on knowledge of the LRDG and understanding of how many drinks it equates to. **Design** Parallel randomized controlled trial. **Setting** On-line, 25 January–1 February 2019. **Participants** Participants (n = 7516) were English, aged over 18 years and drink alcohol. **Interventions** The control group saw existing industry-standard labels; six intervention groups saw designs based on: food labels (serving or serving and container), pictographs (servings or containers), pie charts (servings) or risk gradients. A total of 500 participants (~70 per condition) saw a health warning under the design. **Measurements** Primary outcomes: (i) knowledge: proportion who answered that the LRDG is 14 units; and (ii) understanding: how many servings/containers of beverages one can drink before reaching 14 units (10 questions, average distance from correct answer). **Findings** In the control group, 21.5% knew the LRDG; proportions were higher in intervention groups (all $P < 0.001$). The three best-performing designs had the LRDG in a separate statement, beneath the pictograph container: 51.1% [adjusted odds ratio (aOR) = 3.74, 95% confidence interval (CI) = 3.08–4.54], pictograph serving 48.8% (aOR = 4.11, 95% CI = 3.39–4.99) and pie-chart serving, 47.5% (aOR = 3.57, 95% CI = 2.93–4.34). Participants underestimated how many servings they could drink: control mean = −4.64, standard deviation (SD) = 3.43; intervention groups were more accurate (all $P < 0.001$), best performing was pictograph serving (mean = −0.93, SD = 3.43). Participants overestimated how many containers they could drink: control mean = 0.09, SD = 1.02; intervention groups overestimated even more (all $P < 0.007$), worst-performing was food label serving (mean = 1.10, SD = 1.27). Participants judged the alcohol content of beers more accurately than wine or spirits. The inclusion of a health warning had no statistically significant effect on any measure. **Conclusions** Labels with enhanced pictorial representations of alcohol content improved knowledge and understanding of the UK's low-risk drinking guidelines compared with industry-standard labels; health warnings did not improve knowledge or understanding of low-risk drinking guidelines. Designs that improved knowledge most had the low-risk drinking guidelines in a separate statement located beneath the graphics.

Keywords Alcohol, alcohol unit, cancer, consumer knowledge, graphic labels, health warning label, low-risk drinking guidelines, pictorial labels, product labelling, standard drink.

Correspondence to: Dr Natalie Gold, Public Health England, Wellington House, 133–155 Waterloo Road, London SE1 8UG, UK. E-mail: natalie.gold@phe.gov.uk
Submitted 6 February 2020; initial review completed 23 June 2020; final version accepted 2 November 2020
This article is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

INTRODUCTION

Alcohol consumption is associated with more than 200 diseases, injuries and conditions [1]. For all conditions there is a dose–response relationship—with increasing levels of alcohol consumption there is increasing risk [2]. For some conditions, such as liver disease, this relationship is exponential [3], whereas for other conditions, such as some cancers, it is linear [4]. The most effective way of reducing these risks is reducing individual- and population-level consumption [5]. As such, many governments have developed low-risk drinking guidelines (LRDGs), which commonly include a recommended daily or weekly maximum intake, expressed as numbers of 'standard drinks' or 'units of alcohol' [6,7]. The World Health Organization (WHO) defines a standard drink as 10 g of pure ethanol and advises people not to exceed two standard drinks per day [8]. Although widespread, LRDGs are not universal or uniform. A review of 37 government agency guidelines found that guidelines for low-risk consumption ranged from 10 to 56 g of ethanol per day and that the standard drink sizes (which the guidelines were expressed in) ranged from 8 to 20 g of ethanol, with 10 g as the modal size [7].

The UK government published LRDGs in the 1990s [9,10], which were updated by the UK Chief Medical Officers (CMOs) in 2016 [11]. The weekly drinking guidelines for both men and women state: 'to keep health risks from alcohol to a low level it is safest not to regularly drink more than 14 units a week on a regular basis' [11]. More than 10 million adults in the United Kingdom drink more than the LRDG of 14 units per week [12]. The Department of Health recommended that the CMOs' guidelines be communicated to the general public using visual prompts [13]. In 2011, the Government in England launched the Public Health Responsibility Deal involving voluntary agreements with industry, which included labelling at least 80% of alcohol products with unit content, low-risk guidelines, pregnancy warnings and responsibility statements [14]. However, a market survey conducted in 2014 found that only 57% of labels met best practice as defined by the Portman Group [5].

The LRDG were developed on the principles that: (a) people have a right to accurate information and clear advice concerning alcohol and its health risks and (b) government has a responsibility to ensure that this information is provided for the public in a clear and open way, so that informed choices can be made. However, there is a lack of knowledge and understanding of the LRDG. Recent representative surveys of the adult British population have found that only between 8 and 25% know that the LRDG is 14 units per week [15–17]. Even where people know the guidelines, they may not understand them. The CMOs' guidelines use 'units' of alcohol as a measure. A unit is 10 ml or 8 g of pure alcohol (approximately two teaspoons) [18]. However, research shows that people find it difficult to use units to gauge their alcohol intake, which is not surprising, given that alcoholic drinks vary widely in their strengths and serving sizes [19]. Further, knowledge of the harms that alcohol causes is poor. In a 2018 UK survey, in answer to an open response question concerning which health conditions can result from drinking alcohol, only 40% of respondents identified liver damage/failure as a drinking outcome and 31% reported cancer [16].

A review of the effectiveness of labelling approaches, where labels on alcohol products were enhanced with pictorial representations of alcohol content and health warnings, was carried out to inform this study [Burton et al., unpublished]. The review reported that a range of labelling approaches can effectively increase comprehension of the LRDG and the health risks of alcohol, particularly approaches that use pictorial warnings and messages relating to cancer. The authors concluded that the use of enhanced labels improves comprehension of unit information and the LRDG, especially when labels include information on both these things. It is possible that including both of these components together in alcohol labels can enable a clearer understanding of units and of how many units one can consume within the LRDG. Further, although a growing body of research, including both quantitative and qualitative studies, suggests that adding health warnings to alcohol labels can increase perception of the health risks of alcohol consumption [20–38], no studies have investigated effects on knowledge or understanding of LRDGs of adding health warnings to enhanced labels.

Aims

The main aims of this trial were:

1. to compare the effectiveness of different label designs at conveying knowledge that the LRDG is 14 units; and
2. to compare the effectiveness of different label designs at conveying understanding of how many servings (bottle or can of beer, glass of wine or shot of spirits) or containers (the entire bottle being purchased) could be consumed while remaining within the LRDG.

- to compare the effect of the designs on the perceived risk of alcohol consumption;
- to compare the effect of the designs on the motivation to drink; and
- to compare the effect of the designs on participants' perception of 'health-damaging' drinking (how many units per week they personally thought it would take for a person to 'seriously damage' their health).

Secondary aims were: Finally, we wanted to see whether showing people a health warning alongside our label designs would have a further effect on our secondary outcomes, increasing the perceived risk of alcohol consumption, decreasing the motivation to drink and lowering the level of drinking which people believe to be health-damaging. This was designed as a pilot study, because we were not well-powered; in particular, we could not detect an interaction effect between the warning and the label designs, but we hoped to gain some idea as to whether this hypothesis was worth pursing in future trials.

METHODS

Study design

This was a randomized controlled trial. When participants entered the survey, they were pseudorandomized using computerized random-number generation, which assigned them to one of seven arms (by assigning a number from 1 to 7), each of which saw a different label design. Once it had been determined which of the seven label arms they would be in, a second random-number generation assigned some participants to also see a health warning beneath the design (participants were assigned a new random number between 1 and 100; those who received between one and seven saw labels with the text, and those who received between eight and 100 saw labels without the warning text). See the participant flow in Fig. 1.

Participants did not know the nature of the other interventions. The task was described to them as a 'survey' and they were not told what the other interventions were, or even that other participants might be seeing different labels. Immediately after participants saw the labels, in the same session, they were asked questions to determine their knowledge of the LRDG and their understanding of how much they could drink and stay under the LRDG.

There was an internal study protocol, which can be found in Supporting information, Appendix S1. The study was approved by the Research Support and Governance Office at Public Health England (Ref: R&D 347).

Participants

Participants were recruited from 25 January to 1 February 2019. The trial ended when we had reached the number of responses determined by our power calculations. We recruited participants via a number of third-party panel providers, who have access to a pool of people who have given their consent to be contacted in order to answer on-line questionnaires. Participants were paid a fixed fee of approximately £1 for their time.

Participants were required to be English, aged over 18 years and report drinking alcohol, as measured by the first question of the Alcohol Use Disorders Identification Test (AUDIT-C) questionnaire [23] (see Procedure for full details of the screening). We specified that the sample should be representative of the adult population of England in terms of age, gender and region in which it was (see Supporting information, Table A1, Appendix S2).

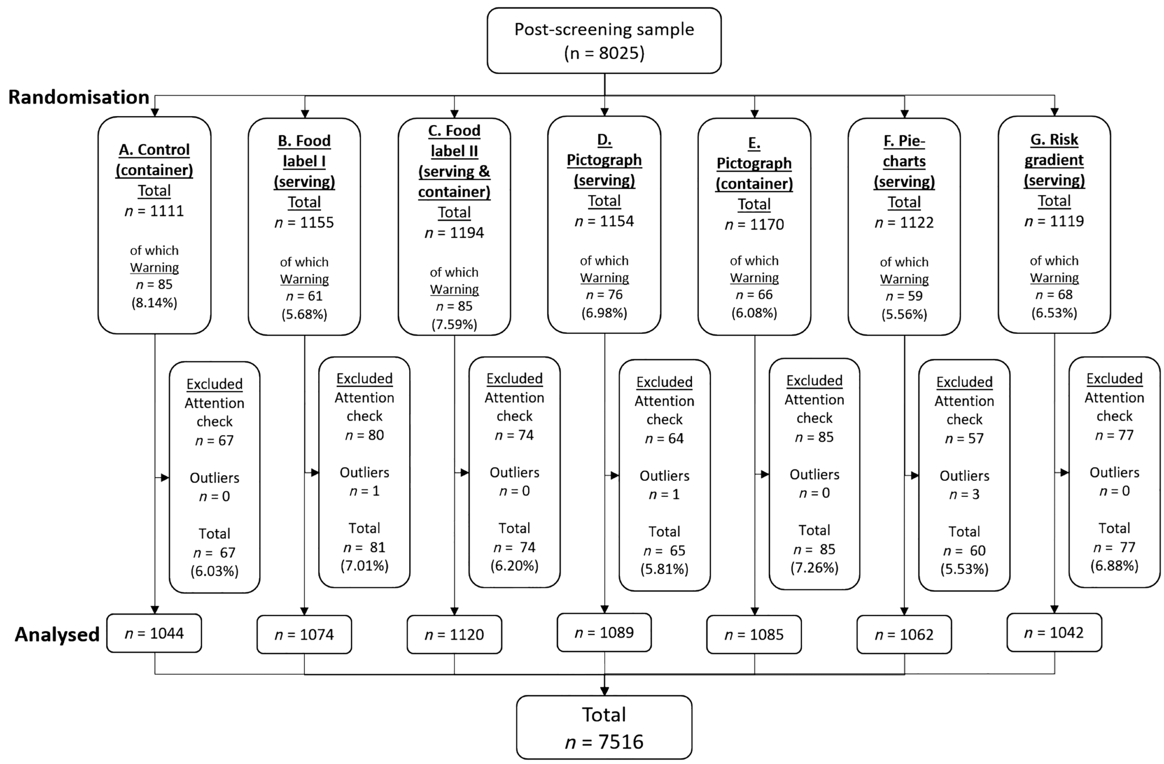

Figure 1 Trial profile.

Interventions

We compared the current industry standard and four other ways of showing information concerning alcohol content: pictograph, pie chart, risk gradient and a design based on food labels. These were taken from current designs in the alcohol and food industries, other designs from the literature and our bespoke pictograph designs. They were among a wider selection of designs that we showed to a focus group with 10 drinkers based in London in December 2018. There were six males and four females; three 18–30-year-olds, five 31–55-year-olds and two 55+-year-olds; representatives from all social classes A, B, C1, C2, D, E; seven white British, three non-white ethnicity; three to four from each of low-income < £25 000, middle-income £25 000–50 000 and high-income > £50 000 groups; four low-risk drinkers, 0–14 units per week, and six increased risk drinkers, 15–35 units for women, 15–50 units for men. We discarded the designs that the focus group participants considered too complicated and used their feedback to refine our preliminary pictograph designs.

We mainly showed the information in terms of servings (for each of the four designs and the control). However, as we were not certain whether showing the information by serving or container would be more effective, we also had one comparison of servings versus containers: pictograph serving and pictograph container held constant in the way that the information was presented (pictograph style), but varied whether it was presented in terms of serving or container. Further, as we wanted to know whether showing both pieces of information would be counterproductive, we had one comparison of single versus multiple framings of information: food label serving and food label serving and container held constant the way that the information was presented (food label style), but allowed us to test the effect of giving participants only servings versus both serving and container information. This gave a total of seven different label designs for alcohol content, including the control.

Participants saw pictures of nine drinks, all seeing the same picture of the bottle and a box with information concerning the ABV and volume of the bottle. Alongside, they saw labels in one of the seven different label designs (see Fig. 2 for examples), as follows.

1. Control (existing industry standard): outline of a bottle with the number of units that are in the entire bottle written inside the outline. No statement of the LRDG.
2. Food label serving: this design was based on food nutrition labels. There was a box that was split into two rows. On the top row was the number of units in a serving, on the bottom row 'x% of the low risk drinking guidelines (14 units per week)'. Above the box, here was a picture of a serving (glass of wine, shot of spirits or bottle of beer) and information concerning the volume of a single serving.
3. Food label serving and container: as for food label serving, but now two boxes, one for servings and one for the container. On the left-hand side was a box showing the number of units in a serving, to exactly the same design as the food label serving, including the picture above. To the right of this box, there was a similar box giving the same information for the container, i.e. the top row of the box had units per container, the bottom row had % of LRDG for the whole container (and repeated the information that the LRDG is 14 units per week), and above the box was a picture of the container and information concerning the volume of alcohol in the container.
4. Pictograph serving: this was a pictograph representation of the proportion of the LRDG that would be consumed in one serving. There was a picture of servings in outline (bottle/can/glass/shots, as appropriate), with the first serving filled in black. The number of servings depicted varied, ranging from five to 26 so that, for example, if one serving was one of five of the LRDG there would be five servings depicted with one filled in, or if one serving was one of 26 of the LRDG there would be 26 servings depicted with one of them filled in. Above the pictograph it said: '1 [serving] = [x] units'. The LRDG was written beneath the pictograph: 'The low-risk drinking guideline is 14 units per week = [y servings]'. The number of servings in this phrase was the same as the number of outline servings in the pictograph.
5. Pictograph container: this was a pictograph representation of the proportion of the LRDG of the whole container's-worth of beverage. There was a picture of containers in outline, filled in black to represent the proportion of a single container/number of containers that would take one up to the LRDG. Above the pictograph it said: '1 bottle = [x] units'. The LRDG was written beneath: 'The low-risk drinking guideline is 14 units per week = [y] bottles'.
6. Pie-chart serving: this was a pie chart that represented the proportion of LRDG in one serving. The number of slices in the pie varied, ranging from five to 26, the number being set so that one serving of the alcohol in question was one slice, so that, for example, if one serving was one of five of the LRDG the pie would be split into five slices with one filled in, or if one serving was one of 26 of the LRDG, the pie would be split into 26 slices with one of them filled in. The LRDG was written beneath: 'The low-risk drinking guideline is 14 units per week = [x servings]'. The number of servings in this phrase was the same as the number of slices in the pie.
7. Risk gradient serving: This had an x-axis in the form of an arrow showing number of units, in colour, fading

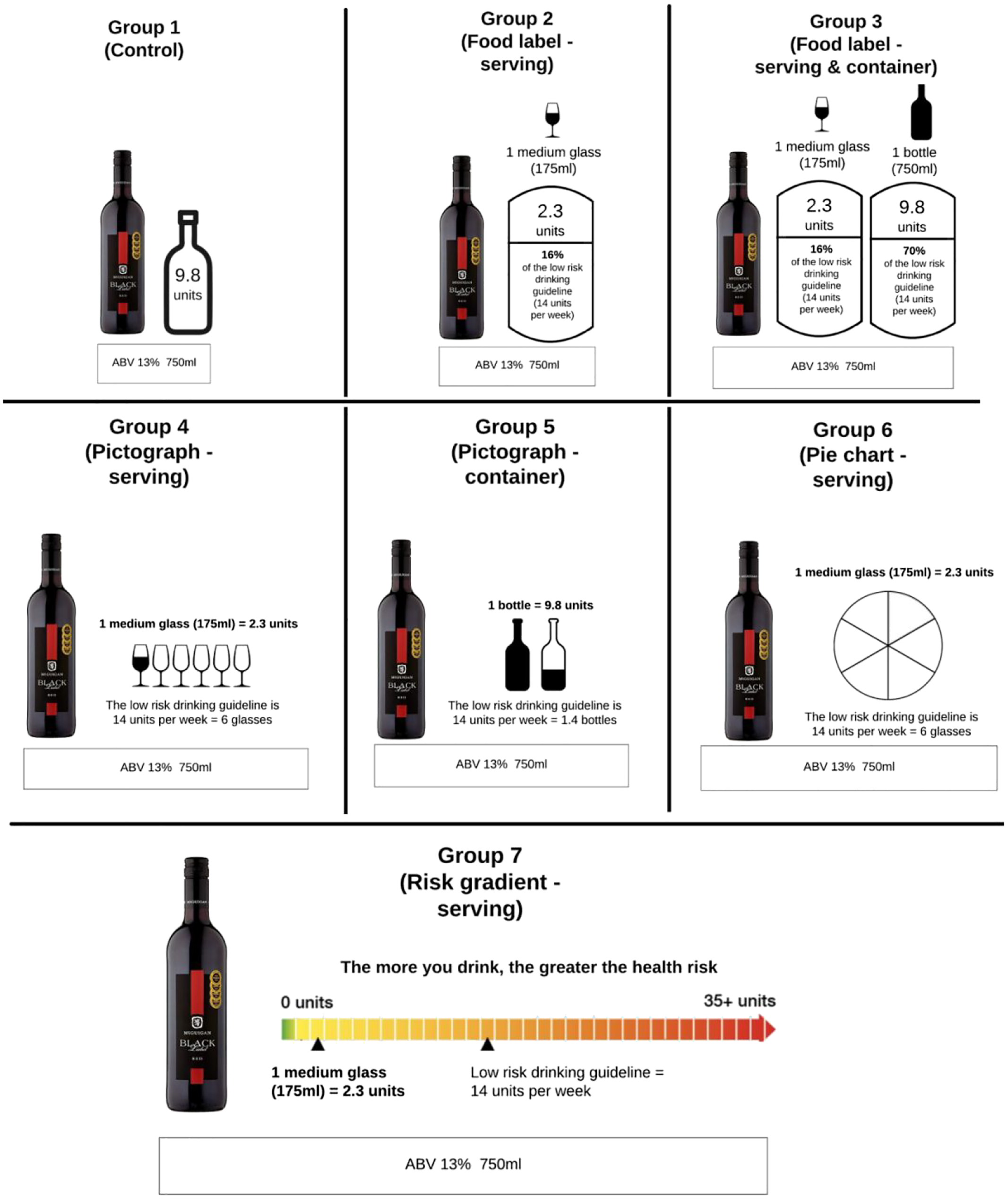

Figure 2 Example of all seven label designs for one of the wines presented. [Colour figure can be viewed at wileyonlinelibrary.com]

from yellow at just above zero, although orange to red at 35, with 'low-risk drinking guideline = 14 units per week' marked at 14 units, which was in the orange part of the spectrum. The number of units in a serving of the beverage was also marked on the axis. 'The more you drink, the greater the health risk' was written above the risk gradient axis.

For our pilot test, 500 participants (~70 in each condition) were randomly assigned to see one of the seven alcohol labels coupled with the text: 'Warning: Alcohol causes cancer' in bold type, with a red line around it beneath the representation of alcohol content (Fig. 3 shows examples of how this appeared in the experiment).

Procedures

Our experiment was conducted on the Behavioural Insight Team's on-line experimentation platform Predictiv.[1] The

[1]Predictiv is an end-to-end platform that aims to make on-line experiments accessible to policymakers and other organizations driven by social impact. The platform provides functionality to run economic experiments and has access to a large international panel, including 200 000 people in the United Kingdom and 1 million in the United States, through a network of on-line panel suppliers. More information can be found on www.predictiv.co.uk.

Figure 3 Example of how the labels with warnings appeared for one beer, one wine and one spirit label. [Colour figure can be viewed at wileyonlinelibrary.com]

full materials are in provided Supporting information, Appendix S2.

Prior to the start of the survey, participants were screened using the first item of the AUDIT-C questionnaire [24]: 'How often do you have a drink containing alcohol?'. Anyone who answered 'never' was excluded from the survey, was not paid and was not counted in the number of participants. Participants who passed the screening test were shown an information statement and asked if they consented to their data being used for research.

Participants were then randomized into one of seven conditions. The conditions were: control (existing industry standard), food label serving, food label serving and container, pictograph serving, pictograph container, pie-chart serving and risk gradient serving. In addition, approximately 70 participants in each condition were randomized to also see a health warning beneath the label. Participants were shown nine pictures of drinks and their ABV, alongside an alcohol label; all nine labels used the design to which they had been allocated (and the warning, if the participant had been allocated to that arm). There were three different beers, three different wines and three different spirits; the drinks were the same for all participants—it was only the labels that changed. See Fig. 2 for examples of the labels. The full set of labels is shown in Supporting information, Appendix S2. Participants could look at labels for as long as they liked and pressed 'next' when they were ready to continue to the questions.

Participants were then asked about the knowledge primary outcome. After answering the knowledge question, participants were explicitly told that the LRDG was 14 units per week, before proceeding to 10 understanding questions, which were presented in a random order. Participants were then asked the secondary outcome questions, followed by some demographic questions. Finally, there was a free text box for feedback.

Measures

Primary outcomes

(1a) Knowledge of the LRDG: 'The government's low-risk drinking guideline recommends that people not regularly drink more than a certain number of alcohol units per week. What do you think the low-risk drinking guideline is?' (free text numerical response).

Our pre-specified primary outcome measure for knowledge of the LRDG was whether participants gave the correct answer (binary variable, coded 1 if participant answered 14 units and 0 otherwise).

(1b) Understanding of the LRDG: We asked 10 understanding questions, which were presented in a random order. The general format of the questions was 'How many [servings/container type (size in ml)] of this [beverage] could you have before reaching 14 units?' (free text numerical response). We grouped the responses into two outcome measures, servings and containers. Note that we considered that a bottle/can of beer was both a serving and a container, so the same two beer questions contributed to both the serving and the container measures.

(2a) Understanding (servings)

There were two questions on each of: (i) beer: 'How many bottles of this beer (330 ml) could you have before reaching 14 units?' and 'How many cans of this beer (586 ml) could you have before reaching 14 units?'; (ii) wine: both 'How many medium-sized glasses of this wine (175 ml) could you have before reaching 14 units?'; and (iii) spirits: both 'How may single shots (25 ml) of this drink could you have before reaching 14 units?'.

For each of the six items we measured distance to the correct response by subtracting the answer given from the correct response (e.g. if the correct answer was 6, then a participant who entered 6 would gain a score of zero, someone who entered 5 would gain a score of −1 and someone who entered 10 would gain a score of 4). Therefore, a positive score represents an overestimation and a negative score represents an underestimation. We then took an average of the six distances to calculate the outcome measure, which is a measure of number of servings from the correct answer.

We also decided to compare participants' understanding of the LRDG measured in terms of units of alcohol, because for health purposes the number of units consumed is what matters. To do this, we converted the distance measure into units of alcohol, i.e. we calculated the number of units each participant was from the correct answer as expressed in units. Again, a positive score represents an overestimation and a negative score represents an underestimation. The score is a measure of the number of units from the correct answer.

(2b) Understanding (containers)

There were two questions on each of: (i) beer: 'How many bottles of this beer (330 ml) could you have before reaching 14 units?' and 'How many cans of this beer (586 ml) could you have before reaching 14 units?'; (ii) wine: both 'How many bottles of this wine (750 ml) could you have before reaching 14 units?'; and (iii) spirits: 'How much of a bottle or whole bottles (700 ml) could you have before reaching 14 units?' and 'How much of a bottle or whole bottles (1 litre) could you have before reaching 14 units?'.

For each of the six items we measured distance to the correct response by subtracting the answer given from the correct response, as detailed for the servings measure 2a, and took an average of the six distances to calculate the outcome measure, measured in number of containers from the correct answer. We also converted the distance measure for containers into units of alcohol, to get the score in terms of the number of units of alcohol from the correct answer, as for the servings measure 2b.

Secondary outcomes

Our secondary outcomes were:

i Perceived personal risk: 'To what extent do you think that cutting down on your drinking would reduce your own risk of alcohol-related disease?' (scale of 1 = not at all likely, 2 = not very likely, 3 = somewhat likely, 4 = quite likely, 5 = extremely likely).

ii Motivation to drink: 'Earlier, you saw the following alcohol label: [beer image no. 3]. To what extent do you agree or disagree with the following statement: This information makes me feel motivated to drink less?' (scale of 1 = strongly disagree 2 = disagree 3 = neither agree nor disagree 4 = agree 5 = strongly agree).

iii Perception of 'damaging' drinking: 'How many units of alcohol do you personally think a person would need to regularly drink per week to seriously damage their health?' (free text numerical response).

Demographics

Participants completed the full AUDIT-C questionnaire, provided demographic information on profession/social grade, smoking status (not presented), ethnicity, highest level of educational attainment (the recruitment companies already had age, gender and in which region of the United Kingdom the participant lives). There was an attention-check question among the demographic items.

Finally, participants were asked for any feedback about the label in an open-text box; for instance, whether they found it useful or confusing, or whether they thought it should be changed.

Statistical analysis

Sample size

A pre-trial power calculation showed that 1000 participants in each arm was sufficient to identify an increase of 4.5–6.4% in the participants who correctly identified the LRDG as 14 units per week, with 80% power and an alpha level of between 0.2 and 5% (we adjusted alpha to account for multiple comparisons using a Hochberg step-up procedure), assuming that 13% of participants in the baseline condition, who saw the existing industry-standard labels, would correctly identify the LRDG as 14 units per week. We also recruited a further 500 participants (approximately 70 in each condition) for a pilot investigation, which included a warning concerning health risks alongside the label.

In order to test knowledge of the LRDG, we ran a logistic regression with whether or not the participant gave the correct answer as the dependent variable, controlling for demographic characteristics, AUDIT-C and warning labels. In order to test understanding, we ran an ordinary least squares (OLS) regression for each of our distance measures (servings and containers), controlling for demographic characteristics, AUDIT-C and warning labels. For our secondary measures we ran OLS regressions, controlling for demographic characteristics, AUDIT-C and warning labels. Data were analysed in Stata version 14.2. The analysis plan was pre-specified in an internal trial protocol (Supporting information, Appendix S1) but it was not pre-registered on a publicly available platform, so the results could be considered exploratory. *Post hoc*, we ran exploratory OLS regressions of our understanding measures disaggregated into different types of alcohol and also with the measures converted into number of units. Upon the request of reviewers, we added a comparison of the proportion in each condition who over- versus underestimated the LRDG, given that they had answered incorrectly.

RESULTS

Participants

We analysed the data of 7516 participants. We excluded 504 participants because they failed the attention check (6.3% of the total 8025 who completed the survey). We excluded a further five participants because their free text numerical response answers were outliers and their survey responses suggested that they had not made a serious attempt to answer the questions (for more detail see Supporting information, Appendix S1). The participant flow is shown in Fig. 1.

Participants were recruited via a number of panel providers. Eligible participants were English, aged over 18 years and drinkers of alcohol. There were 3798 women and participants were aged between 18 and 99 years [mean = 44.15, standard deviation (SD) = 16.45]. Details of our participants' baseline characteristics can be found in Table 1. Our sample was recruited to be representative of the adult population of England in terms of age, gender and region.

Primary outcome: knowledge of LRDG

More participants under- than overestimated the LRDG and the distribution was skewed (see Fig. 5): the modal response was the correct answer of 14, the median was 12 and the interquartile range was 9 (from 5 to 14).

In the control group, only 21.3% of participants correctly answered that the LRDG was 14 units per week. A logistic regression showed that participants in all the intervention conditions had a more accurate knowledge of the LRDG than those in the control condition (all $P < 0.001$, summary statistics and adjusted odds ratios (aORs) are reported in Table 2). There appears to be a cluster of three best-performing designs (pictograph container, 51.1%, followed by pictograph serving 48.8% and pie-chart serving, 47.5%—the three that had the LRDG in a separate statement, beneath the graphics) and three that did not perform quite so well, even though they performed better than the control (food label serving 38.7%, risk gradient serving, 35.6% and food label serving and container, 32.9%), as shown in Fig. 4, where the unadjusted 95% CIs do not overlap between the two clusters or the control.

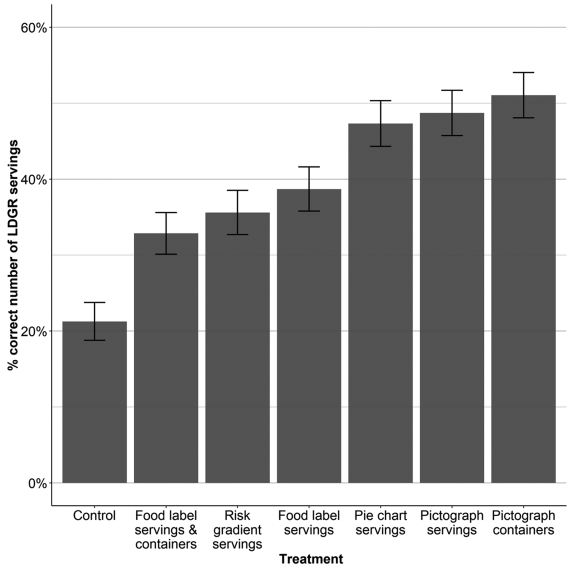

Figure 4 Bar chart low-risk drinking guidelines (LRDG) knowledge (%) correct with 95% confidence interval (CI) bars (by condition).

More than 80% of those who provided the incorrect LRDG gave an answer that was less than the LRDG; this was true in all conditions (see Fig. 6). Although the proportion who answered the LRDG incorrectly varied depending on the label design, given that participants had answered wrongly, there were no statistically significant differences as to whether they were likely to under- or overestimate between conditions, $\chi^2_{(6)} = 10.22$, $P = 0.11$.

Several of the variables that we controlled for in the OLS regression were related to knowledge. Those who were aged 55+ were more likely to answer the question correctly than 18–24-year-olds, and people with any level of education from secondary upwards were more accurate than people with no secondary education. Lower social

Table 1 Baseline demographics characteristics of the seven trial arms and overall for the whole trial.

Trial arm	*Number in trial arm*	*Number (%) of females*	*Age Mean (SD)*	*AUDIT-C score Mean (SD)*
Control	1044	516 (50.6%)	44.18 (16.75)	4.96 (2.67)
Food label (servings)	1074	558 (52.0%)	43.58 (16.23)	5.04 (2.69)
Food label (servings and containers)	1120	569 (50.8%)	44.05 (16.34)	5.00 (2.70)
Pictograph (containers)	1085	571 (52.6%)	43.94 (16.35)	5.09 (2.71)
Pictograph (servings)	1089	543 (49.9%)	43.94 (16.56)	5.17 (2.75)
Pie chart (servings)	1062	525 (49.4%)	44.15 (16.31)	5.09 (2.78)
Risk gradient (servings)	1042	516 (49.5%)	45.26 (16.61)	5.03 (2.73)
Overall	7516	3798 (50.5%)	44.15 (16.45)	5.06 (2.72)

AUDIT-C = Alcohol Use Disorders Identification Test; SD = standard deviation.

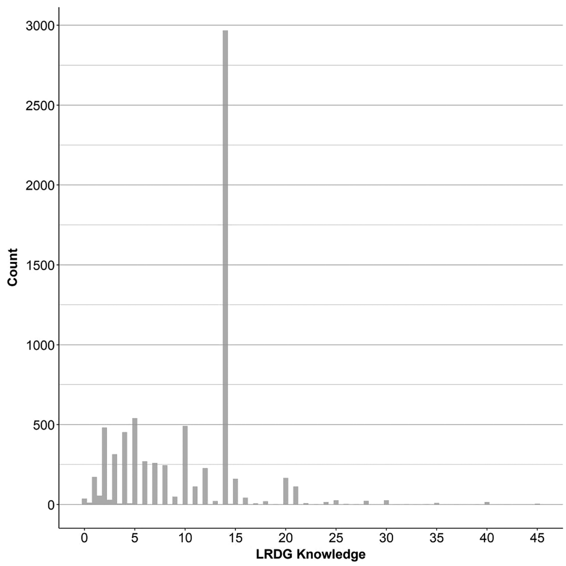

Figure 5 Distribution of participant responses to low-risk drinking guidelines (LRDG) knowledge (LRDG = 14) excluding outlier responses above the 99th percentile.

grades (C2DE) answered less accurately than higher grades (ABC1); black, Asian and mixed-race ethnicities answered less accurately than white. There were regional variations. There was no statistically significant relationship between answers to the knowledge measure and sex, AUDIT-C score or having seen the warning.

Primary outcome: understanding of LRDG servings

The understanding (servings) measure had good internal consistency (Cronbach's alpha = 0.67).

Every group underestimated how many servings it takes to reach 14 units (see Table 3). Control group participants were the least accurate on our primary outcome measure (the average of their distance measures for two beers, two wines and two spirits); their estimates were furthest from the LRDG of 14 units (mean servings = – 4.64, SD = 3.43; mean units = – 4.43, SD = 3.95). The best-performing group was pictograph serving (mean servings = – 0.93, SD = 3.43; mean units =– 0.96, SD = 2.46). An OLS regression showed that participants in all the interventions had a clearer understanding of how many servings they could consume and remain under the 14-unit LRDG than those in the control condition [all $P < 0.001$, see Table 4 for full model and confidence intervals (CIs)]. Comparing the four intervention designs that only gave information in terms of servings, risk gradient serving performed the worst—it did not have overlapping CIs with any of the other three for accuracy of number of servings in the adjusted model—and the numerical ordering of performance was pictograph serving > pie-chart serving > food label serving > risk gradient serving. There was no evidence of any detriment in understanding of LRDG servings from adding container information to the food label design: food label serving had overlapping CIs with food label serving and container in the adjusted model. It is notable that the pictograph container condition, the only intervention not to give information in servings, while more accurate than the control, was less accurate than all the other intervention arms (no overlapping 95% CIs, either adjusted or unadjusted). There was no effect of having seen a warning label.

The inaccuracy was driven by the estimates for wines, and especially spirits. Figure 7 shows the understanding estimates for servings, disaggregated into wine, beer and spirts. For beer, all intervention groups gave similar and accurate answers to questions concerning how many servings they could have. Within each label design the CIs of the understanding estimates for servings of beer, alcohol and spirits do not overlap, with participants being least accurate regarding servings of spirits. When the estimates are expressed in terms of units the numerical ordering is preserved, but some of the CIs overlap (see Fig. 7).

Primary outcome: understanding of LRDG containers

The understanding (containers) measure had good internal consistency (Cronbach's alpha = 0.66).

Table 2 Knowledge of the low-risk drinking guidelines (LRDG): proportion of participants who correctly identified the LRDG as 14 units and adjusted odds ratios (aORs) from a binary logistic regression controlling for demographics; ordered from smallest to largest aOR.

Trial arm	*Number of participants in the trial arm*	*Number correctly identifying LRDG*	*% Correctly identifying LRDG*	*aOR*	*95% CIs*		*P-value*
Control	1044	222	21.3	–	–	–	–
Food label Servings and Containers	1120	368	32.9	1.85	1.52	2.26	< 0.001
Risk gradient	1042	371	35.6	2.09	1.71	2.55	< 0.001
Food label Serving	1074	416	38.7	2.44	2.01	2.97	< 0.001
Pie chart	1062	504	47.5	3.57	2.93	4.34	< 0.001
Pictograph Serving	1089	531	48.8	4.11	3.39	4.99	< 0.001
Pictograph Container	1085	554	51.1	3.74	3.08	4.54	< 0.001

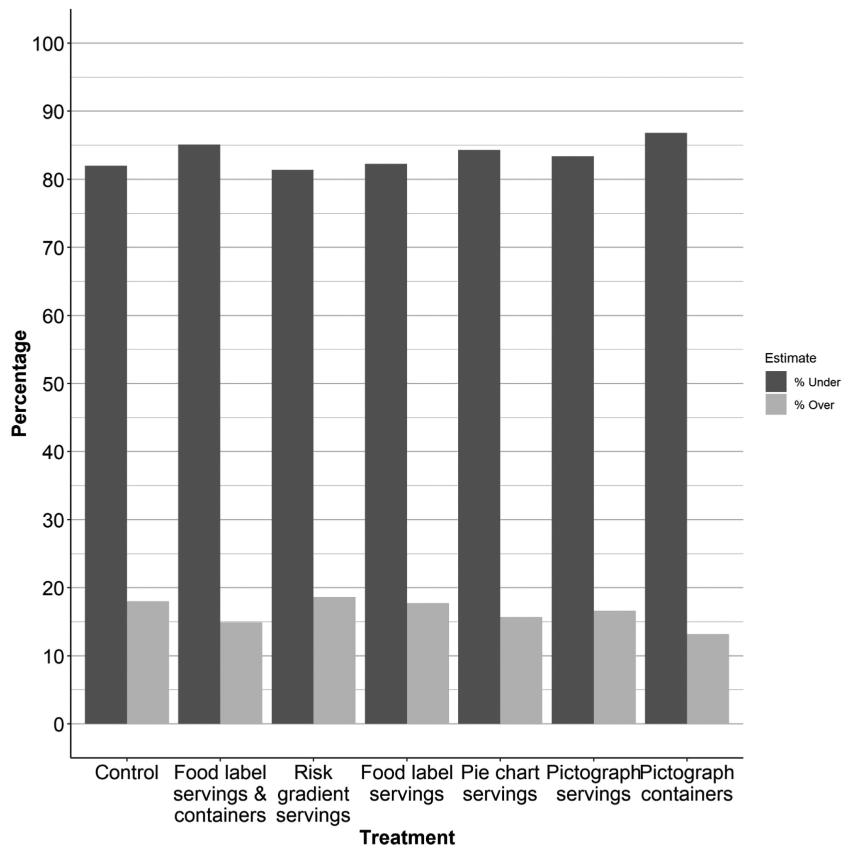

Figure 6 Participants who gave the incorrect answer to the low-risk drinking guidelines (LRDG), percentage of those who were wrong who under-*versus* over-estimated.

Table 3 Understanding of the low-risk drinking guidelines (LRDG): distance from the correct answer for questions concerning how many servings/containers could be consumed before reaching 14 units (each measure is an average of six answers: two beer, two wine and two spirits).

	Accuracy of understanding (servings)					
	Servings			*Units*		
Trial arm (ordered most to least accurate)	*Mean (SD)*	*95% CIs*		*Mean (SD)*	*95% CIs*	
Pictograph serving	−0.93 (2.17)	−1.06	−0.80	−0.96 (2.46)	−1.10	−0.81
Pie-chart serving	−1.11 (2.49)	−1.26	−0.96	−1.12 (2.93)	−1.30	−0.94
Food label serving	−1.21 (2.75)	−1.37	−1.04	−1.20 (3.02)	−1.38	−1.02
Food label serving and container	−1.40 (2.85)	−1.56	−1.23	−1.36 (3.10)	−1.54	−1.18
Risk gradient serving	−1.84 (3.63)	−2.06	−1.62	−1.61 (4.91)	−1.91	−1.31
Pictograph container	−3.45 (3.44)	−3.66	−3.25	−2.96 (4.20)	−3.21	−2.71
Control	−4.64 (3.43)	−4.85	−4.44	−4.43 (3.95)	−4.67	−4.19
Trial arm	**Accuracy of understanding (containers)**					
	Containers			**Units**		
(ordered most to least accurate)	**Mean (SD)**	**95% CIs**		**Mean (SD)**	**95% CIs**	
Control	0.09 (1.02)	0.03	0.16	6.00 (14.08)	5.14	6.85
Pictograph container	0.22 (0.99)	0.16	0.27	6.44 (15.21)	5.54	7.35
Food label serving and container	0.40 (1.09)	0.33	0.46	8.31 (15.44)	7.41	9.22
Pie-chart serving	0.80 (1.17)	0.73	0.87	14.81 (18.47)	13.70	15.92
Risk gradient serving	0.81 (1.56)	0.72	0.91	15.74 (20.14)	14.51	16.96
Pictograph serving	0.90 (1.13)	0.84	0.97	15.78 (18.63)	14.68	16.89
Food label serving	1.10 (1.27)	1.02	1.17	19.62 (20.36)	18.40	20.84

CI = confidence interval; SD = standard deviation.

Table 4 Understanding of the low-risk drinking guidelines (LRDG): ordinary least squares (OLS) regression with accuracy of estimate of how many servings/containers could be drunk and the drinker still remain under the 14-unit per week LRDG.

	Servings: distance to correct answer[a] (compared to baseline category for categorical variables)				*Containers: distance to correct answer[b] (compared to baseline category for categorical variables)*			
Characteristic	*β (SE)*	*95% CIs*		*P-value*	*β (SE)*	*95% CIs*		*P-value*
Treatment (baseline = control)								
Food label serving	3.42 (0.13)	3.16	3.67	< 0.001	1.02 (0.05)	0.92	1.12	< 0.001
Food label serving and container	3.24 (0.13)	2.99	3.49	< 0.001	0.32 (0.05)	0.22	0.42	< 0.001
Pictograph serving	3.70 (0.13)	3.44	3.95	< 0.001	0.82 (0.05)	0.72	0.92	< 0.001
Pictograph container	1.17 (0.13)	0.92	1.43	< 0.001	0.14 (0.05)	0.04	0.24	0.007
Pie-chart serving	3.53 (0.13)	3.27	3.78	< 0.001	0.72 (0.05)	0.62	0.82	< 0.001
Risk gradient serving	2.79 (0.13)	2.54	3.05	< 0.001	0.74 (0.05)	0.64	0.84	< 0.001
Age, years (baseline = 18–24)								
25–54	−0.09 (0.11)	−0.29	0.12	0.42	−0.30 (0.4)	−0.38	−0.21	< 0.001
55+	0.42 (0.11)	0.20	0.64	< 0.001	−0.33 (0.5)	−0.42	−0.24	< 0.001
Female (baseline = male)	0.2 (0.7)	−0.12	0.16	0.75	−0.09 (0.03)	−0.15	−0.04	0.001
Social grade C2DE (baseline = ABC1)	−0.27 (0.07)	−0.41	−0.12	< 0.001	−0.02 (0.03)	−0.08	0.04	0.47
Ethnicity (baseline = white)								
Black	−0.62 (0.23)	−1.07	−0.18	0.006	0.44 (0.09)	0.26	0.62	< 0.001
Asian	−0.57 (0.19)	−0.94	−0.21	0.002	0.34 (0.07)	0.19	0.48	< 0.001
Mixed	−0.65 (0.24)	−1.12	−0.17	0.007	0.23 (0.10)	0.04	0.41	0.018
Other	−0.11 (0.40)	−0.90	0.68	0.78	0.35 (0.16)	0.04	0.67	0.027
Region (baseline = North)								
South and East	0.22 (0.09)	0.05	0.38	0.012	−0.03 (0.03)	−0.10	0.04	0.35
Midlands	−0.15 (0.10)	−0.35	0.04	0.12	−0.07 (0.04)	−0.14	0.01	0.097
London	−0.40 (0.11)	−0.61	−0.16	0.001	0.05 (0.05)	−0.03	0.14	0.23
AUDIT-C (numerical, 1–12)	0.02 (0.01)	−0.01	0.04	0.20	0.00 (0.01)	−0.01	0.01	0.997
Highest education (baseline = none)								
Secondary	0.67 (0.26)	0.16	1.18	0.01	−0.05 (0.10)	−0.25	0.15	0.62
Post-secondary/vocational	1.21 (0.26)	0.70	1.71	< 0.001	−0.12 (0.10)	−0.32	0.08	0.26
Undergraduate or higher	1.67 (0.26)	0.76	1.78	< 0.001	−0.21 (0.10)	−0.42	−0.01	0.038
Warning (baseline = no warning)	0.05 (0.14)	−0.23	0.32	0.75	0.04 (0.05)	−0.06	0.15	0.42
Constant	−5.73 (0.30)	−6.33	−5.14	< 0.001	0.52 (0.12)	0.29	0.76	< 0.001
R^2	0.18				0.10			
Sample size	7481				7500			

AUDIT-C = Alcohol Use Disorders Identification Test; SE = standard error; CI confidence interval. [a]The ‘Servings’ outcome was measured by taking the average of people’s estimates of how many beers (two questions), servings of wines (two questions) and servings of spirits (two questions) it takes to reach 14 units, and then subtracting the technically correct answer from this. The analysis excludes 35 participants who gave ineligible responses for at least one of these six questions. [b]The ‘Containers’ outcome was measured by taking the average of people’s estimates for how many beers (two questions), containers of wines (two questions) and containers of spirits (two questions) it takes to reach 14 units, and then subtracting the technically correct answer from this. The analysis excludes 16 participants who gave ineligible responses for at least one of these six questions.

Every group overestimated how many containers it takes to reach 14 units (see Table 3). Control participants were the most accurate when averaging across their estimates for how many containers of beer, wine and spirits they could have; their estimates were closest to the 14-unit LRDG (mean containers = 0.09, SD = 1.02; mean units = 6.00, SD = 14.08). Food label serving was the numerically worst-performing group (mean containers = 1.10, SD = 1.27; mean units = 19.62, SD = 20.36). An OLS regression showed that participants in all the interventions had a worse understanding of how many containers they could consume and remain under the LRDG than those in the control condition (all $P < 0.001$; see Table 4 for the full model, including CIs). The most accurate two intervention conditions were pictograph container and food labels servings and container, the two that gave information in terms of containers, which had 95% CIs that did not overlap with any other of the other interventions (although the adjusted CIs overlapped with each other). Comparing the four designs that only gave information in terms of servings, the numerical ordering of performance is pie-chart serving > risk gradient serving > pictograph serving > food label serving, although there were overlaps in the 95% CIs of the coefficients in the adjusted model. There was no effect of having seen a warning label.

Again, these inaccuracies were driven almost entirely by participants’ estimates for wine and spirits. The beer

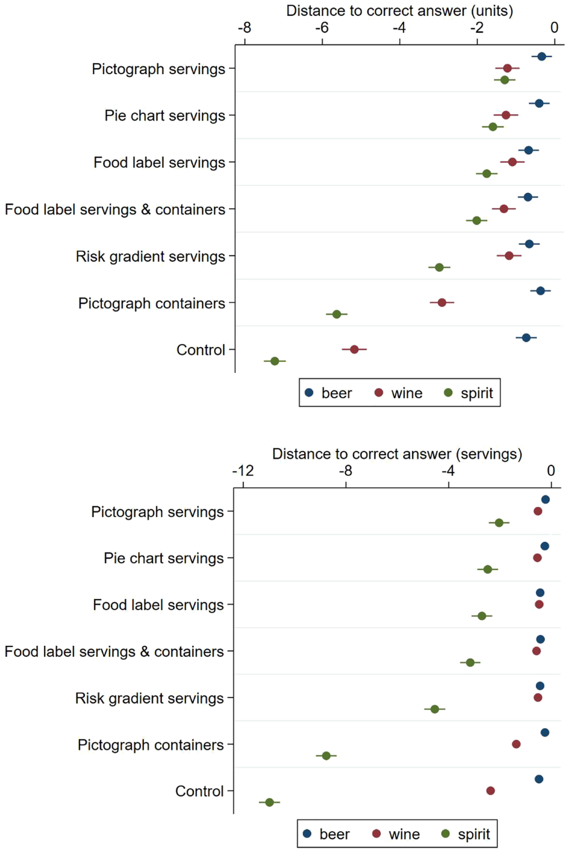

Figure 7 Understanding of the low-risk drinking guidelines (LRDG) (servings): how many servings of alcohol can be consumed while remaining under the LRDG? Mean distance from the correct answer in (a) servings and (b) units, ordered from most to least accurate (in terms of aggregate average measure), showing 95% confidence intervals (CIs) from an ordinary least squares (OLS) regression controlling for demographics. [Colour figure can be viewed at wileyonlinelibrary.com]

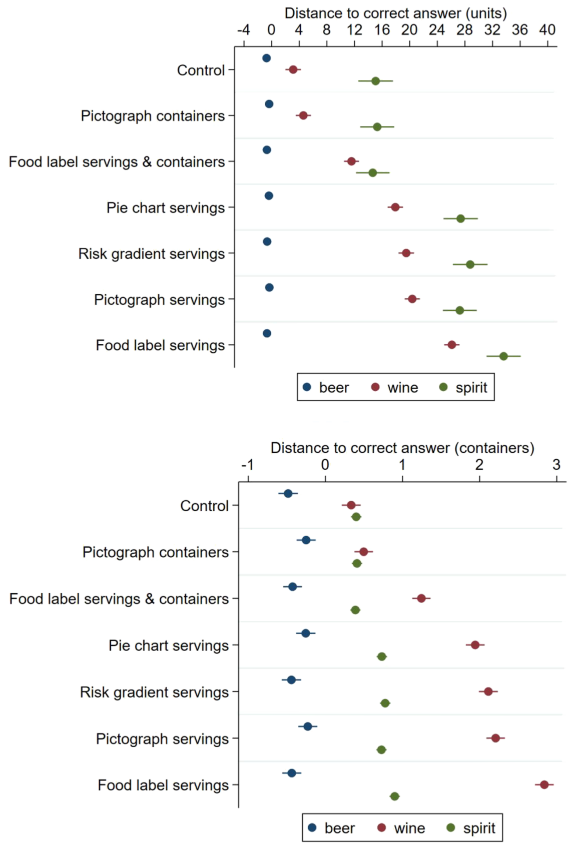

Figure 8 Understanding of the low-risk drinking guidelines (LRDG) (containers): how many servings of alcohol can be consumed while remaining under the LRDG? Mean distance from the correct answer in (a) containers and (b) units, ordered from most to least accurate (in terms of aggregate average measure), showing 95% CIs from an ordinary least squares (OLS) regression controlling for demographics. [Colour figure can be viewed at wileyonlinelibrary.com]

estimates were most accurate in all conditions and, within each condition, the CIs for beer estimates did not overlap with those for wine or spirits (see Fig. 8). When estimates were expressed in terms of containers, the wine estimates were numerically most inaccurate and CIs did not overlap with spirits estimates in any condition except the control and pictograph containers. When estimates were expressed in terms of units, then spirits estimates were numerically most inaccurate and the CIs did not overlap with wine for any condition, apart from food servings and containers.

Secondary outcomes

For our secondary measures, we found that participants in all conditions on average thought that it was 'quite likely' that cutting down on their alcohol consumption would reduce the risk of disease (mean = 3.88, SD = 1.22); on average, they neither agreed nor disagreed that the alcohol label made them less motivated to drink (mean = 3.23, SD = 1.03), and the average estimate of how many units per week a person would need to drink to seriously damage their health was 24 units (mean = 26.24, SD = 62.60) (see Table 5 for a complete breakdown by trial arm). We ran OLS regressions on the secondary measures and found that the enhanced label designs had no effect on the perceived personal risk of drinking or on the perception of health-damaging drinking, but they all decreased the stated motivation to drink compared to the control, albeit by a very small amount (0.1–0.3 points on a five-point scale). There was no effect of the warnings. For the full models see Table 6.

DISCUSSION

All our enhanced alcohol-label designs improved knowledge of the LRDG. In the control group, only 21.5% of participants correctly answered that the LRDG was 14 units per week, but knowledge was higher in every intervention arm (proportion of correct answers ranged from 32.9 to 51.5%). Our enhanced designs improved understanding

Table 5 Secondary outcomes [ordinary least squares (OLS) regressions]: perceived personal risk of own drinking (1–5), motivation to drink (1–5) and subjective perception of high-risk drinking (numerical free text response).

	Perceived risk[a]			*Motivation to drink*[b]			*Perception of health-damaging drinking*[c]		
Trial arm	*Mean (SD)*	*95% CIs*		*Mean (SD)*	*95% CIs*		*Mean (SD)*	*95% CIs*	
Control	3.87 (1.16)	3.80	3.94	3.07 (1.08)	3.00	3.13	25.00 (36.50)	22.78	27.22
Food label serving	3.84 (1.13)	3.77	3.90	3.21 (1.04)	3.14	3.27	26.02 (46.93)	23.21	28.83
Food label serving and container	3.89 (1.11)	3.83	3.96	3.23 (1.03)	3.17	3.29	24.88 (23.51)	23.51	26.26
Pictograph serving	3.89 (1.10)	3.83	3.96	3.23 (1.04)	3.16	3.29	25.30 (21.02)	24.05	26.55
Pictograph container	3.87 (1.11)	3.80	3.93	3.33 (1.00)	3.27	3.39	26.22 (48.91)	23.30	29.13
Pie-chart serving	3.90 (1.11)	3.83	3.96	3.29 (0.99)	3.23	3.35	26.03 (25.69)	24.48	27.57
Risk gradient serving	3.91 (1.12)	3.85	3.98	3.27 (1.05)	3.20	3.33	23.90 (17.11)	22.86	24.94
Overall average	3.88 (1.12)	3.86	3.91	3.23 (1.03)	3.21	3.26	25.34 (33.54)	24.58	26.10

[a]To what extent do you think that cutting down on your drinking would reduce your own risk of alcohol related disease? From 1 (not at all likely) to 5 (extremely likely). [b]Earlier, you saw the following alcohol label: [beer image no. 3]. To what extent do you agree or disagree with the following statement: 'This information makes me feel motivated to drink less' [from 1 (strongly disagree) to 5 (strongly agree)]. [c]How many units of alcohol do you personally think a person would need to regularly drink per week to seriously damage their health? Free text numerical response.

of the LRDG when it was expressed in terms of servings, but decreased understanding when it was expressed in terms of containers. The enhanced designs had no effect on the perceived personal risk of drinking or on the subjective perception of high-risk drinking, but they all decreased the stated motivation to drink compared to the control, albeit by a very small amount. The addition of a cancer warning had no effect on any of our measures.

It is not surprising that our interventions increased the level of knowledge of the LRDG, as the existing industry-standard label was the only one that did not explicitly state the LRDG. The 21.5% who responded with the correct LRDG in the control condition is consistent with the results of recent UK surveys, where the proportion of participants correctly reporting the LRDG has varied from 8 to 25% [15–17]. Of our new enhanced designs, pictograph servings, pictograph container and pie chart faired particularly well, with 47–51% of participants correctly reporting the LRDG. In all three of these designs, the LRDG was given in a separate statement, beneath the graphics, which may have made it particularly salient. Participants who gave an incorrect answer were more likely to under- than overestimate the LRDG: in all conditions, even as the number giving an accurate answer increased, more than 80% of those who were incorrect gave an underestimate.

Participants in all seven conditions underestimated the number of servings that they could drink and still remain under the LRDG. Understanding in terms of servings was more accurate in the intervention groups. This replicates the findings of two previous studies [25,26]. Conversely, participants in all conditions overestimated the number of containers of alcohol that they could drink and still remain under the LRDG, and accuracy of understanding decreased in the intervention conditions. When we disaggregated in terms of type of alcohol, we found that participants' estimates for beer were similar and fairly accurate across conditions; the inaccuracies and differences were driven by estimates for wine, and especially spirits (when denominated in terms of number of units). One reason why our participants were more accurate for beer may be because people often drink entire containers (bottles or cans) of beer as a single serving, but they usually need to pour out servings of wine and spirits, and they rarely drink a whole container of spirits in one sitting. This suggests that, potentially, we could improve understanding of the alcohol content of wine and spirits if containers and serving vessels had lines indicating standard units, so that people are more aware of the number of servings they are consuming.

The most effective labels differed depending on whether understanding was measured in terms of units or of containers. Unsurprisingly, the accuracy of understanding estimates varied depending on whether the design participants had seen was congruent with the question: the designs showing servings led to more accurate answers to questions concerning servings, while designs showing containers led to more accurate estimates of containers. Pictographs were highly successful when the presentation and the question were congruent, but among the least successful labels when they were not congruent. Pie-chart servings was a reasonable performer on both understanding measures. Interestingly, adding container information to the food label design, as well as servings, increased understanding of containers without any detriment in understanding of servings. However, we cannot infer that providing both types of information on the other designs, which had more reliance upon graphics, would have the same effect, although this might be worthy of further investigation.

From a public health perspective, when deciding whether to present information in terms of servings or

Table 6 Secondary outcomes [ordinary least squares (OLS) regressions]: perceived personal risk of own drinking (1–5), motivation to drink (1–5) and subjective perception of high-risk drinking (numerical free text response).

	Perceived risk[a]				*Motivation to drink*[b]				*Perception of health-damaging drinking*[c]			
Characteristic	*β (SE)*	*95% CIs*		*P-value*	*β (SE)*	*95% CIs*		*P-value*	*β (SE)*	*95% CIs*		*P-value*
Treatment (baseline = control)												
Food label serving	−0.04 (0.05)	−0.13	−0.06	0.42	0.14 (0.04)	0.06	0.23	0.001	0.98 (1.44)	−1.85	3.81	0.50
Food label serving and container	0.02 (0.05)	−0.07	0.12	0.64	0.17 (0.04)	0.08	0.26	< 0.001	−0.20 (1.43)	−2.99	2.60	0.89
Pictograph serving	0.02 (0.05)	−0.07	0.12	0.65	0.17 (0.04)	0.08	0.26	< 0.001	−0.04 (1.44)	−2.86	2.77	0.98
Pictograph container	−0.01 (0.05)	−0.10	0.09	0.89	0.28 (0.04)	0.19	0.36	< 0.001	0.91 (1.44)	−1.91	3.72	0.53
Pie-chart serving	0.03 (0.05)	−0.07	0.12	0.57	0.23 (0.04)	0.14	0.32	< 0.001	0.93 (1.45)	−2.11	3.57	0.61
Risk gradient serving	0.05 (0.05)	−0.05	0.15	0.31	0.210 (0.04)	0.12	0.30	< 0.001	−1.31 (1.46)	−4.16	1.53	0.37
Age, years (baseline = 18–24)												
25–54	−0.02 (0.04)	−0.10	0.06	0.67	0.0 (0.04)	−0.07	0.07	0.99	3.01 (1.18)	0.69	5.33	0.011
55+	−0.13 (0.04)	−0.21	−0.04	0.003	−0.09 (0.04)	−0.17	−0.02	0.02	5.34 (1.27)	2.85	7.82	<0.001
Female (baseline = male)	0.14 (0.03)	0.09	0.19	<0.001	0.04 (0.2)	−0.00	0.09	0.08	0.83 (0.78)	−0.70	2.36	0.29
Social grade C2DE (baseline = ABC1)	−0.02 (0.03)	−0.08	0.03	0.39	−0.01 (0.03)	−0.06	0.04	0.60	−0.05 (0.83)	−1.68	1.58	0.95
Ethnicity (baseline = white)												
Black	0.17 (0.09)	0.00	0.34	0.044	0.21 (0.08)	0.06	0.36	0.53	4.33 (2.53)	−0.63	9.28	0.09
Asian	0.11 (0.07)	−0.03	0.25	0.12	0.24 (0.06)	0.11	0.36	< 0.001	−3.14 (2.07)	−7.19	0.91	0.13
Mixed	−0.10 (0.9)	−0.27	0.08	0.29	0.05 (0.08)	−0.11	0.21	0.007	−0.17 (2.68)	−5.41	5.08	0.95
Other	−0.5 (0.15)	−0.35	0.24	0.72	0.14 (0.14)	−0.13	0.41	0.32	1.53 (4.50)	−7.30	10.36	0.73
Region (baseline = North)												
South and East	−0.01 (0.03)	−0.08	0.05	0.65	−0.05 (0.03)	−0.11	0.01	0.10	0.60 (0.96)	−1.27	2.48	0.53
Midlands	0.02 (0.04)	−0.06	0.09	0.63	−0.0 (0.03)	−0.07	0.06	0.93	1.39 (1.11)	−0.79	3.57	0.21
London	−0.01 (0.04)	−0.09	0.08	0.88	0.08 (0.04)	−0.05	0.16	0.04	0.75 (1.28)	−1.76	3.26	0.56
AUDIT-C (numerical, 1–12)	−0.00 (0.0)	−0.01	0.01	0.63	−0.04 (0.00)	−0.05	−0.03	< 0.001	1.89 (0.14)	1.61	2.18	<0.001
Highest education (baseline = none)												
Secondary	0.27 (0.10)	0.08	0.46	0.006	0.03 (0.09)	−0.14	0.21	0.73	3.44 (2.90)	−2.24	9.12	0.24
Post-secondary/ vocational	0.28 (0.10)	0.09	0.46	0.004	−0.01 (0.09)	−0.18	0.16	0.91	2.35 (2.87)	−3.27	7.98	0.41
Undergraduate or higher	0.28 (0.10)	0.09	0.47	0.004	0.04 (0.09)	−0.14	0.21	0.67	1.14 (2.90)	−4.55	6.82	0.70
Warning (baseline = no warning)	0.00 (0.05)	−0.10	0.11	0.93	0.07 (0.05)	−0.02	0.17	0.12	−1.28 (1.54)	−4.29	1.74	0.41
Constant	3.60 (0.11)	3.38	3.82	<0.001	3.26 (0.10)	3.05	3.46	< 0.001	9.18 (3.38)	2.56	15.80	0.01
R^2	0.01				0.03				0.03			
Sample size	7516				7516				7516			

AUDIT-C = Alcohol Use Disorders Identification Test; SE = standard error; CI confidence interval. [a]To what extent do you think that cutting down on your drinking would reduce your own risk of alcohol-related disease? [from 1 (not at all likely) to 5 (extremely likely)]. [b]Earlier, you saw the following alcohol label: [beer image no. 3]. To what extent do you agree or disagree with the following statement: 'This information makes me feel motivated to drink less' [from 1 (strongly disagree) to 5 (strongly agree)] [c]How many units of alcohol do you personally think a person would need to regularly drink per week to seriously damage their health? Free text numerical response.

containers it would be better to choose whichever keeps health risks lower. There are two considerations, which pull in opposite directions. In order to encourage lower alcohol consumption, it is better if participants underestimate (rather than overestimate) how much they can drink, which suggests contextualizing how much alcohol it takes to reach the LRDG in terms of servings. However, total alcohol intake will depend not only upon how many servings people think they can have, but also on whether they can accurately track how many servings they are drinking. People tend to have difficulty pouring standard drinks, with overpouring being the norm [19]. Recent studies have found that when people are asked to pour out a 'normal serving' of spirits, on average they pour 2 units of alcohol, similar to a 50-ml double shot [27,28]. Even if people underestimate the amount of servings they are allowed, if they overpour their drinks (overestimating the size of a standard serving) then labels that are denominated in terms of servings may lead to higher alcohol consumption than labels that are denominated in terms of containers.

The health warning did not affect responses to either the primary or secondary outcome measures. It seems likely that this was despite participants noticing it, as the warning was large and in a red box, and both size and colour have been shown to be important regarding whether or not people pay attention to warnings [29,30]. Although other studies have found that cancer warnings increase the perceived risk of drinking alcohol [20], reduce the stated motivation to drink [25] and reduce the stated future drinking intentions [21], in those studies the cancer warnings were always being compared to other types of warning. Two other studies presented participants with warning labels compared to a control condition with no label and both found no effect of the text warning compared to the control [22,31]. Pictures and graphical warnings have been found to be more effective than text warnings [22,32]. Further, it may not be surprising that a warning alone has no effect, as there is a large body of evidence on 'fear appeals', which shows that fear-control processes can interfere with the motivation to take precautions [33] and that fear appeals are only effective in the presence of high self-efficacy for taking action to prevent the risk [34,35]; where there is low self-efficacy, fear appeals may lead to avoidance or reactance [34].

We found that enhanced labels can improve knowledge and understanding of the LRDG, but they did not affect our secondary outcomes: the perceived risk of alcohol consumption, the motivation to drink and the level of drinking which people believe to be health-damaging. In general, our participants tended to underestimate both the LRDG and of the number of servings they could drink and remain beneath it, which may be protective. Although our results suggest that improving alcohol labelling alone is unlikely to change behaviour, enhanced labels could still facilitate informed choice. This raises the prospect that improving knowledge and understanding might lead to an increase in alcohol consumption, if people adjust consumption upwards to reach the LRDG. Therefore, it is important that people understand the nature of the dose–response relationship, whereby risk increases with drinking, rather than regarding the LRDG as a threshold for safe drinking.

We randomized a large number of individuals to each condition, which is a strength of our trial. The main limitation of our trial is that we ran an on-line experiment and our results may not generalize well to field settings, including supermarkets, which are the places where people are most likely to see a label of the type we tested on a bottle before they buy it. In our on-line setting the labels were presented on a screen, and although the size of the labels on mobile screens were comparable to the size of labels on a bottle, participants could zoom in if they wished. Labels that display a large amount of information (e.g. the food label serving and container design) or which are wide in design (e.g. the risk gradient) may perform worse in real life if they need to be shrunk to fit onto standard alcohol packaging. Furthermore, although the average participant in our experiment spent approximately 60 seconds reviewing the various example labels, we know from laboratory studies that people pay little attention to alcohol health warnings or responsible drinking statements [29,36], and we expect that they would pay even less attention to them at point of sale. Lastly, although our sample was designed to be representative of basic population characteristics, our participants were a self-selecting group who had agreed to be on a panel and answer questions for money. Potentially, their behaviour may not be representative of the average member of the population. Therefore, although our study shows that our labels would improve knowledge and understanding if people pay attention to them, we cannot be sure to what extent those results would generalize to a field setting where people might not pay attention.

This study was about comprehension of risk. Although we asked about intention to reduce alcohol consumption, not only did we not find a meaningful effect of the labels, we also know that there is an intention–behaviour gap: stated intentions may not translate into behaviour [37]. As well as testing comprehension at point of sale and investigating how to encourage people to pay more attention to labels in the field, future research could investigate whether different label designs have any effect on purchasing and consumption behaviour.

Taken together, these results show that improved pictorial designs to communicate alcohol risk can lead to better knowledge and understanding of LRDGs. All our custom designs improved knowledge that the UK LRDG is 14 units, compared to industry-standard labels. Designs that had the LRDG in a separate statement, beneath the graphics, improved knowledge the most. For understanding, different designs performed best depending on whether the question was how many servings could be consumed while remaining under the LRDG or how many containers (and the safe number of servings was underestimated, so improving understanding could increase the amount that people think it is safe to drink). However, the results suggest that there is room for improvement in existing alcohol labels.

Declaration of interests

None.

Acknowledgements

This study was supported by Public Health England. We thank the Bristol Tobacco and Alcohol Research Group for input on the label designs, Lucy Porter for designing the CONSORT flowchart, Adam Winter on commissioning and conception of the design and Jeric Kison for comments on the paper and supporting the project management. F.G.'s research is supported by the National Institute for Health Research Applied Research Collaboration Northwest London and the National Institute for Health Research School of Public Health Research. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

AUTHOR CONTRIBUTIONS

Natalie Gold: Conceptualization; formal analysis; methodology; supervision. **Mark Egan:** Data curation; formal analysis; investigation; methodology; validation. **Kristina Londakova:** Methodology; resources. **Abigail Mottershaw:** Data curation; investigation; methodology; software; supervision. **Hugo Harper:** Project administration; supervision. **Robyn Burton:** Conceptualization. **Clive Henn:** Conceptualization; funding acquisition; methodology; supervision. **Maria Smolar:** Conceptualization; funding acquisition; methodology; project administration. **Matthew Walmsley:** Conceptualization; funding acquisition; methodology; supervision. **Rohan Arambepola:** Formal analysis; visualization. **Robin Watson:** Formal analysis; visualization. **Sarah Bowen:** Formal analysis; visualization. **Felix Greaves:** Conceptualization; funding acquisition; methodology; supervision.

References

1. World Health Organization (WHO). Global Status Report on Alcohol and Health. Geneva, Switzerland: WHO; 2018.
2. Babor T., Caetano R., Casswell S., Edwards G., Giesbrecht N., Graham K., *et al. Alcohol: No Ordinary Commodity: Research and Public Policy New York*. Oxford: Oxford University Press; 2010.
3. Rehm J., Taylor B., Mohapatra S., Irving H., Baliunas D., Patra J., *et al.* Alcohol as a risk factor for liver cirrhosis: a systematic review and meta-analysis. *Drug Alcohol Rev* 2010; **29**: 437–45.
4. Bagnardi V., Rota M., Botteri E., Tramacere I., Islami F., Fedirko V., *et al.* Alcohol consumption and site-specific cancer risk: a comprehensive dose–response meta-analysis. *Br J Cancer* 2015; **112**: 580–93.
5. Burton R., Henn C., Lavoie D., O'Connor R., Perkins C., Sweeney K., *et al.* A rapid evidence review of the effectiveness and cost-effectiveness of alcohol control policies: an English perspective. *Lancet* 2017; **389**: 1558–80.
6. Furtwængler N. A., de Visser R. O. Lack of international consensus in low-risk drinking guidelines. *Drug Alcohol Rev* 2013; **32**: 11–8.
7. Kalinowski A., Humphreys K. Governmental standard drink definitions and low-risk alcohol consumption guidelines in 37 countries. *Addiction* 2016; **111**: 1293–8.
8. Babor T., Higgins-Biddle J. *Brief Intervention for Hazardous and Harmful Drinking: A Manual for use in Primary Care; 2001*. Geneva, Switzerland: World Health Organization; 2011.
9. Lord President of the Council *Action Against Alcohol Misuse*. London: HMSO; 1991.
10. Department of Health *The Health of the Nation—A Strategy for Health in England*. London, UK: HMSO; 1992.
11. Department of Health. UK Chief Medical Officers' Low Risk Drinking Guidelines. London, UK: Department of Health; 2016.
12. Health and Social Care Information Centre. Health Survey for England 2014: chapter 8. Adult Alcohol Consumption. Leeds, UK: Health and Social Care Information Centre; 2015.
13. Department of Health. Communicating the UK Chief Medical Officers' Low Risk Drinking Guidelines. London, UK: Department of Health; 2017.
14. Department of Health. Public Health Responsibility Deal: Alcohol Pledges. London, UK: Department of Health; 2011.
15. Rosenberg G., Bauld L., Hooper L., Buykx P., Holmes J., Vohra J. New national alcohol guidelines in the UK: public awareness, understanding and behavioural intentions. *J Public Health* 2017; **40**: 549–56.
16. Alcohol Health Alliance UK. How We Drink, What We Think. Public Views on Alcohol and Alcohol Policies in the UK. London, UK: Alcohol Health Alliance; 2018.
17. Buykx P., Li J., Gavens L., Hooper L., Gomes de Matos E., Holmes J. Self-reported knowledge, correct knowledge and use of UK drinking guidelines among a representative sample of the English population. *Alcohol Alcohol* 2018; **53**: 453–60.
18. Department of Health. Alcohol Units: a Brief Guide. London, UK: Department of Health; 2008.
19. Kerr W. C., Stockwell T. Understanding standard drinks and drinking guidelines. *Drug Alcohol Rev* 2012; **31**: 200–5.
20. Jongenelis M. I., Pratt I. S., Slevin T., Chikritzhs T., Liang W., Pettigrew S. The effect of chronic disease warning statements on alcohol-related health beliefs and consumption intentions among at-risk drinkers. *Health Educ Res* 2018; **33**: 351–60.
21. Miller E. R., Ramsey I. J., Baratiny G. Y., Olver I. N. Message on a bottle: are alcohol warning labels about cancer appropriate? *BMC Public Health* 2016; **16**: 139.
22. Pettigrew S., Jongenelis M., Chikritzhs T., Slevin T., Pratt I. S., Glance D., *et al.* Developing cancer warning statements for alcoholic beverages. *BMC Public Health* 2014; **14**: 786.
23. Wigg S., Stafford L. D. Health warnings on alcoholic beverages: perceptions of the health risks and intentions towards alcohol consumption. *PLOS ONE* 2016; **11**: e0153027.
24. Bush K., Kivlahan D. R., McDonell M. B., Fihn S. D., Bradley K. A. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. *Arch Intern Med* 1998; **158**: 1789–95.
25. National Institute for Health Research. COVID-19 and Mental Health Studies Register. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684826/Alcohol_use_disorders_identification_test_for_consumption__AUDIT_C_.pdf

26. Blackwell A. K., Drax K., Attwood A. S., Munafò M. R., Maynard O. M. Informing drinkers: can current UK alcohol labels be improved? *Drug Alcohol Depend* 2018; **192**: 163–70.
27. Hobin E., Vallance K., Zuo F., Stockwell T., Rosella L., Simniceanu A., *et al*. Testing the efficacy of alcohol labels with standard drink information and national drinking guidelines on consumers' ability to estimate alcohol consumption. *Alcohol Alcohol* 2017; **53**: 3–11.
28. Gill J. S., Donaghy M. Variation in the alcohol content of a 'drink' of wine and spirit poured by a sample of the Scottish population. *Health Educ Res* 2004; **19**: 485–91.
29. Boniface S., Kneale J., Shelton N. Actual and perceived units of alcohol in a self-defined 'usual glass' of alcoholic drinks in England. *Alcohol Clin Exp Res* 2013; **37**: 978–83.
30. Pham C., Rundle-Thiele S., Parkinson J., Li S. Alcohol warning label awareness and attention: a multi-method study. *Alcohol Alcohol* 2017; **53**: 39–45.
31. Al-Hamdani M., Smith S. M. Alcohol warning label perceptions: do warning sizes and plain packaging matter? *J Stud Alcohol Drugs* 2016; **78**: 79–87.
32. Krischler M., Glock S. Alcohol warning labels formulated as questions change alcohol-related outcome expectancies: a pilot study. *Addict Res Theory* 2015; **23**: 343–9.
33. Chen Y., Yang Z. J. Message formats, numeracy, risk perceptions of alcohol-attributable cancer, and intentions for binge drinking among college students. *J Drug Educ* 2015; **45**: 37–55.
34. Ruiter R. A., Abraham C., Kok G. Scary warnings and rational precautions: a review of the psychology of fear appeals. *Psychol Health* 2001; **16**: 613–30.
35. Witte K., Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. *Health Educ Behav* 2000; **27**: 591–615.
36. Tannenbaum M. B., Hepler J., Zimmerman R. S., Saul L., Jacobs S., Wilson K., *et al*. Appealing to fear: a meta-analysis of fear appeal effectiveness and theories. *Psychol Bull* 2015; **141**: 1178–204.
37. Kersbergen I., Field M. Visual attention to alcohol cues and responsible drinking statements within alcohol advertisements and public health campaigns: relationships with drinking intentions and alcohol consumption in the laboratory. *Psychol Addict Behav* 2017; **31**: 435–46.
38. Sheeran P., Webb T. L. The intention–behavior gap. *Soc Personal Psychol Compass* 2016; **10**: 503–18.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Trial protocol and data analysis plan*

Appendix S2. Experimental design

