Do alcohol product labels stating lower strength verbal description, percentage alcohol-by-volume, or their combination affect wine consumption? A bar laboratory adaptive randomised controlled trial

Milica Vasiljevic[1,2] (iD), Daniel Frings[3], Mark Pilling[1] & Theresa M. Marteau[1] (iD)

Behaviour and Health Research Unit, University of Cambridge, Cambridge, UK,[1] Department of Psychology, Durham University, Durham, UK[2] and Centre for Addictive Behaviours Research, London South Bank University, London, UK[3]

ABSTRACT

Background and Aims   A previous research study concluded that wine and beer labelled as lower in strength increase consumption compared with the same drinks labelled as regular strength. The label included both a verbal and numerical descriptor of strength. The present study aimed to estimate the effect of each of these label components. **Design**   Adaptive, parallel group randomised controlled trial, comprising an internal pilot sample ($n_1 = 90$) and a confirmatory sample ($n_2 = 57$). **Setting**   University bar laboratory in London, United Kingdom (UK). **Participants**   A total of 147 weekly wine drinkers were sampled from a nationally representative English panel. **Intervention**   Participants were randomised to one of three groups to taste test wine in a bar-laboratory, varying only in the label displayed: (i) verbal descriptor only (Super Low); (ii) numerical descriptor only (4% alcohol by volume (ABV)); and (iii) verbal descriptor and numerical descriptor combined (Super Low 4%ABV) (each group $n = 49$). **Measurements**   The primary outcome was total volume (ml) of wine consumed. **Findings**   Participants randomised to the numerical descriptor label group (4%ABV: $M = 155.12$ ml, $B = 20.30$; 95% CI = 3.92, 36.69; P value = 0.016) and combined verbal and numerical descriptor label group (Super Low 4%ABV: $M = 154.59$ ml, $B = 20.68$; 95% CI = 4.32, 37.04; P value = 0.014) drank significantly greater amounts than those randomised to the verbal descriptor label group (Super Low: $M = 125.65$ ml). **Conclusions**   This bar laboratory study estimated that a greater quantity of 'lower' strength wine was consumed when the label included a numerical strength descriptor compared with a verbal only strength descriptor.

Keywords   Alcohol consumption, alcohol labelling, bar lab, lower strength alcohol labelling, public health, randomised controlled trial.

Correspondence to: Milica Vasiljevic, Department of Psychology, Upper Mountjoy, South Road, Durham University, Durham DH1 3LE, UK. E-mail: milica.vasiljevic@durham.ac.uk;

Theresa M. Marteau, Behaviour and Health Research Unit, Forvie Site, Robinson Way, University of Cambridge, Cambridge CB2 0SR, UK. E-mail: theresa.marteau@medschl.cam.ac.uk

INTRODUCTION

Worldwide, 5.3% of all deaths and 5.1% of the global burden of disease can be attributed to alcohol consumption [1]. Policy changes encouraging the greater availability and sale of lower strength alcoholic drinks have been mooted as having the potential to reduce total alcohol consumption and associated harms at the population level [2]. In the United Kingdom (UK), industry representatives have called for extending the range of lower strength alcohol labelling above the current cap of 1.2% alcohol by volume

(%ABV) [3]. For lower strength alcoholic drinks to achieve their full potential for reduced consumption at the population level, two conditions need to be met: (i) the occasions during which alcohol is consumed must not increase (potentially extending the total time during which alcohol is consumed; [4,5]); and (ii) consumers must not compensate for the lower strength of these drinks by consuming more (thereby resulting in higher overall alcohol consumption; [6]).

Although evidence on this topic is scant, the few studies conducted to date suggest that extending the range of

lower strength alcohol labelling may lead to paradoxical effects. Labels displaying alcohol units seem to be counterintuitively used as a reference cue to identify and purchase stronger or cheapest-for-strength alcohol products, highlighting a possible negative effect of more prominent labelling of the alcohol content of drinks [7,8]. More recent studies found that the marketing materials used both by producers and retailers in the United Kingdom for lower strength alcoholic drinks suggested extending the occasions suitable for alcohol consumption [9]. In another study, weekly wine and beer consumers sampled from the UK population mirrored such claims, by reporting that they perceived lower strength alcoholic drinks as suitable for consumption on more occasions and by more varied consumer groups when compared to drinks of regular strength [10]. Furthermore, although weekly drinkers' self-reported understanding of the alcohol content of lower strength alcoholic drinks was superior to knowledge of content of regular strength alcoholic drinks, this better understanding did not translate in less harmful consumption [11]. Specifically, participants in a bar laboratory setting consumed approximately 20% more wine or beer when it was labelled as lower in strength.

However, Vasiljevic and colleagues could not examine whether the effects of the lower strength alcohol labelling stemmed from the verbal or the numerical descriptor of strength because all the labels denoting lower alcohol strength featured both verbal (e.g. Super Low Alcohol) and numerical information (e.g. 4%ABV) of alcohol strength [11]. This is an important consideration for any change in legislation pertaining to labelling and promotion of lower strength alcoholic drinks because of the potential for adverse effects arising from such labelling at the population level. The present study aimed to fill this gap. We examined whether; the verbal descriptor (Super Low) or the percentage alcohol by volume (4%ABV), or their combination would impact consumption. As no prior evidence exists that could inform directional hypotheses, exploratory analyses were undertaken to test the effects of these labels.

METHODS

Trial design

A between-subjects parallel group randomised controlled trial with one independent factor of three levels corresponding to the label that accompanies wine for consumption: (i) label displaying only a verbal descriptor (Super Low); (ii) label displaying only a numerical (%ABV) descriptor (4%ABV); and (iii) label displaying a combination of both a verbal and a numerical (%ABV) descriptor (Super Low 4%ABV).

Because prior studies did not provide an estimate of effect size for the comparisons between the three experimental groups the study incorporated an internal pilot

allowing us to estimate the required total sample size in an adaptive fashion [12–14]. In the internal pilot, the initial sample size of 90 participants (30 per group) was chosen as a pragmatic number to estimate the likely effect size of the comparison between the three groups, with a view to estimating a required total sample size to detect a likely effect [13]. Before commencement, it was pre-specified and prospectively pre-registered that if the required total sample size for the full trial was estimated to be <300 following analysis of the internal pilot data (i.e. futility criteria), we would recruit the remaining sample and analyse the data from all participants together [14] (see also our trial registration https://doi.org/10.1186/ISRCTN33451258). After analysing the data from the internal pilot (Table 1), it was estimated from simulations that a full trial would require the recruitment of 57 additional participants (19 per group) to attain 93% power to detect a difference between the group randomised to a verbal descriptor only label (Super Low) versus the group randomised to a combination label of both a verbal and numerical descriptor (Super Low 4%ABV); and 82% power to detect a difference between the group randomised to a verbal descriptor only label (Super Low) versus the group randomised to a numerical descriptor only label (4%ABV). Because this was within the pre-specified range in our trial registration, we therefore recruited the additional 57 participants, resulting in a total of 147 participants for the full trial.

Participants

In total, 147 weekly alcohol drinkers (age 18+) with a stated preference for wine were recruited via a research agency (https://www.icmunlimited.com/) from a panel representative of the general population of England (with interlocking quotas set for age, gender and socio-economic status [SES]). Exclusion criteria included pregnancy (women only), current medication use and a history of neurological or psychiatric disorders. Once eligibility for the study was ascertained, participants were randomised to one of the three groups varying in the labels used to describe the wines they were invited to taste, but not in the actual wines (see Procedure). See Table 2 for the characteristics of the sample recruited.

Table 1 Summary statistics (ml) of the experimental groups at the interim analysis (*n* = 30 per experimental group).

	Mean (SD)	*95% CI*
Super Low	112.47 (57.47)	91.01 – 133.93
4% ABV	163.10 (90.25)	129.40 – 196.80
Super Low 4%ABV	160.70 (70.99)	134.19 – 187.21

Table 2 Participant demographic and drinking characteristics.

	Super Low	*4%ABV*	*Super Low 4%ABV*
Sample size	49 (33.3)	49 (33.3)	49 (33.3)
Gender			
Male	26 (53.1)	22 (44.9)	24 (49)
Female	23 (46.9)	27 (55.1)	25 (51)
Age group (years)			
18–44	25 (51)	24 (49)	24 (49)
45–70	24 (49)	25 (51)	25 (51)
Social grade			
Low (DE)	16 (32.7)	16 (32.7)	16 (32.7)
Medium (C1C2)	17 (34.7)	17 (34.7)	17 (34.7)
High (AB)	16 (32.7)	16 (32.7)	16 (32.7)
Ethnicity			
White	37 (75.5)	27 (55.1)	30 (61.2)
Other	12 (24.5)	22 (44.9)	19 (38.8)
Education[a]			
Up to 4 GCSEs	5 (10.2)	5 (10.2)	2 (4.1)
1 A-level	7 (14.3)	6 (12.2)	0
2+ A levels	12 (24.5)	9 (18.4)	11 (22.4)
University	24 (49)	29 (59.2)	34 (69.4)
Missing/NA	1 (2)	0	2 (4.2)
Income (£)[b]			
[0,15.5 K] p.a.	2 (4.1)	3 (6.1)	4 (8.2)
[15.51 K,25.5 K] p.a.	1 (2)	1 (2)	6 (12.2)
[25.51 K,40 K] p.a.	17 (34.7)	13 (26.5)	11 (22.4)
[> 40.01 K] p.a.	26 (53.1)	30 (61.2)	25 (51)
Missing/NA	3 (6.1)	2 (4.1)	3 (6.1)
Index of multiple deprivation (IMD)[c]			
Quintile1	9 (18.4)	12 (24.5)	5 (10.2)
Quintile2	11 (22.4)	20 (40.8)	18 (36.7)
Quintile3	10 (20.4)	6 (12.2)	12 (24.5)
Quintile4	10 (20.4)	4 (8.2)	5 (10.2)
Quintile5	3 (6.1)	5 (10.2)	2 (4.1)
Missing/NA	6 (12.2)	2 (4.1)	7 (14.3)
Risky drinking (AUDIT-C)			
Mean (SD)	4.8 (1.68)	5.2 (1.79)	4.97 (1.69)
Taste test duration (minutes)			
Mean (SD)	7.85 (1.33)	7.77 (1.41)	7.78 (1.39)

Percentages appear in parentheses. [a]GCSEs (General Certificate of Secondary Education) are usually taken at age 15–16 in the UK; A-Levels at age 17–18. [b]Income bands are expressed per annum. [c]Index of multiple deprivation (IMD) denotes neighbourhood-level deprivation; Quintile 1 reflects the highest level of deprivation and Quintile 5 the lowest level of deprivation.

Intervention

To avoid possible ceiling effects, participants in each experimental group were presented with three identical glasses filled with equal amounts (125 ml per glass) of wine (actual %ABV ~5.5%). Because the study used a between-subjects design, a cover story was used purporting that the three glasses contain three samples of the same wine manufactured by the same producer and with the same ingredients, but fermented in vessels made from different materials that can result in variations in taste. Therefore, the participants were told they will partake in a taste-preference task and rate the three wine samples. The labels comprised small pieces of card placed in front of the glasses according to randomisation. The taste-preference task is a validated method for assessing alcohol consumption in laboratory settings and is validated as an analogue for participants' real-world alcohol use outside of the lab [15].

A glass containing 250 ml of water was available at all times as a palate cleanser. To reinforce the cover story, participants were asked to rate how pleasant, strong tasting, sweet and fizzy the wines were (adapted from [16]). Participants were told they could drink as much or as little as they liked to make their ratings and were informed that the taste test would last up to 10 minutes ($M = 7.80$, $SD = 1.37$). The online Supporting Information

contains detailed information on the intervention and taste-preference task set-up (see Figure S1 - S3).

Measures

Primary outcome

Total volume of drink consumed (in ml) was the primary outcome. High precision scales (Smart Weigh Model PL11B) were used to measure the total volume of drink consumed during the taste test.

Secondary outcome

Product appeal was measured by two items: 'How likely are you to buy/drink this wine?' (answered on scales ranging from 1 = very unlikely to 7 = very likely) (Spearman's $\rho = 0.78$, $P < 0.001$).

Other measures

Risky drinking was assessed using the alcohol use disorder identification text-consumption (AUDIT-C) [17]. A sample item asked 'How often do you have six or more drinks on one occasion?' with responses ranging from never, less than monthly, monthly, weekly, daily or almost daily. Responses to the three items were summed and dichotomised to denote riskier (scoring above 5) versus less risky drinking patterns (scoring below 5) [18].

Motivation to reduce consumption within the next 6 months were gauged via three items; 'Thinking about the next 6 months: I intend/want/will try to drink less alcohol'. Responses were recorded on 7-point scales anchored from 1 (strongly disagree) to 7 (strongly agree) (Cronbach's $\alpha = 0.93$).

Self-licensing was measured using two items: 'If I were to have a low alcohol drink, I would feel like I deserved to have something stronger for my next drink'; and 'If I were to have a low alcohol drink, I would feel like I could have more than my usual number of drinks'. Participants responded using 7-point scales ranging from 1 (strongly disagree) to 7 (strongly agree) (Spearman's $\rho = 0.36$, $P < 0.001$).

Demographic characteristics including age, gender, ethnicity, and SES (highest educational qualification, income, occupation and index of multiple deprivation [19]) were recorded.

Procedure

The study was approved by the University of Cambridge Psychology Research Ethics Committee (PRE.2017.095) and the London South Bank University School of Applied Sciences Research Ethics Committee (SAS1802). The study was conducted in a bar laboratory mimicking a 'bar' environment, based in a central London location. The internal pilot was carried out in the period 21 May 2018 to 13 June 2018 and the second phase 9 July 2018 to 2 August 2018. Participants were randomly allocated to one of the three experimental groups using random permuted blocks stratified by age (18–44 years, 45–80 years), gender and occupational SES (low, medium, high) constructed before the commencement of the study by a statistician (M.P.) blinded to the groups. This was performed in R 3.4.1 by creating a list of all possible combinations, and then randomly removing some possibilities until the required list length was achieved [20]. The flow of participants through the study can be seen in the CONSORT flow diagram below (Fig. 1).
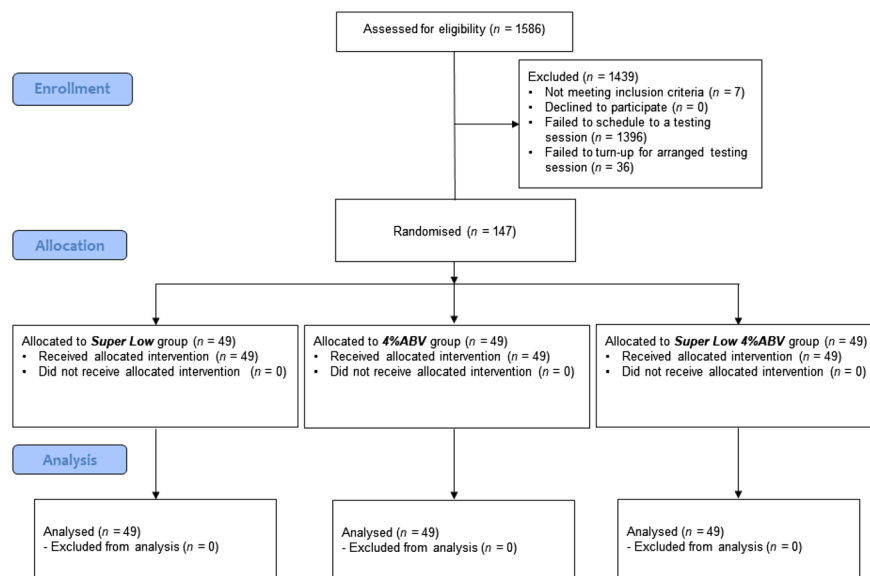
Figure I CONSORT diagram of participant flow through the study

The randomisation allocation to experimental group was concealed from the market research agency recruiters who assigned participants to a unique participation number according to age, gender and SES occupational status. Participants were blinded to assignment of experimental group (open-ended questions at the end of the testing session confirmed that participants were not aware of the study aims).

On arrival, participants were told they were undertaking a taste-preference task in which they would rate the quality of a new wine developed for the market and provided with the cover story outlined above. Participants were then provided with the three glasses of wine and undertook the taste test. Following this, participants completed a survey containing the secondary outcome and other measures in the order they are described above. Participants were then probed about their understanding of the aims of the study and debriefed about the nature of the study and told its true purpose. At this point, participants underwent another breathalyser test to gauge their intoxication. Participants who were above the driving limit were asked to either stay in the laboratory until the effects of the alcohol have dissipated, or take public transportation when leaving the testing venue. Once participants vacated the bar laboratory, the fluid they did not consume was measured to ascertain how much wine they had drank.

Analysis

Analyses were conducted using SPSS 25 [21]. The internal pilot analysis ($n_1 = 90$) was based on the pre-specified stopping boundary from the usual O'Brien–Fleming spending function, the trial would have stopped if significance had been achieved at the $P \leq 0.0007$ level for either of the two pairwise comparisons, which did not occur. At the interim, it was decided that as the effect sizes were very similar for the two contrasts between the Super Low experimental group and the other two experimental groups, the pre-planned primary analysis should be powered for two comparisons with the Super Low experimental group, resulting in two pairwise comparisons.

Examination of the distribution of the data revealed that responses to the primary outcome were not normally distributed (Shapiro–Wilk test), which could not be corrected by transformation, therefore the rank of total volume consumed was used in an extension of the non-parametric Kruskal–Wallis test. Although the secondary outcome was also not normally distributed, regression diagnostics were satisfactory when a transformation was used (i.e. log [8-score]), which corrected the positive skew.

The subsequent analyses were therefore carried out using regressions that included the stratification covariates [22,23], which all had good diagnostics. The threshold for significance set using the O'Brien–Fleming boundary for the final stage analysis was set at $P \leq 0.0472$ to adjust for the interim analysis. Because there were two pairwise comparisons of interest, the threshold for significance of each comparison was set at $P \leq 0.0236$. Stratification covariates were reasonably balanced between the experimental groups. Below, we present descriptive statistics in both non-parametric and parametric formats to aid understanding.

RESULTS

Primary outcome

Wine consumption differed significantly between experimental groups. Participants who tasted wine with the label combining verbal and numerical descriptors (Super Low 4%ABV: Mean Rank = 80.96, M = 154.59 ml; 95%CI 134.55–174.64 ml, SD = 69.79) drank significantly more ($P = 0.014$) when compared to those participants who tasted the wine labelled with a verbal descriptor only (Super Low: Mean Rank = 61.23, M = 125.65 ml; 95%CI 104.78–146.53 ml, SD = 72.67). Participants who tasted wine labelled with the numerical descriptor only (4%ABV: Mean Rank = 79.81, M = 155.12 ml; 95%CI 132.20–178.06 ml, SD = 79.81) drank significantly more ($P = 0.016$) when compared to those participants who tasted wine labelled with only a verbal descriptor (Super Low) (see Table 3 for the full model). For a graphical presentation of consumption levels across the three experimental groups, see Fig. 2.

Extending the primary models to include moderating variables (see online Supporting Information Table S2 for syntax and inferential statistics); gender, age and SES occupation yielded no statistically conclusive interactions (with the exception of a main effect of gender, whereby men drank significantly more than women, but this effect of gender did not differ between experimental groups). Similarly, when examining main effects and interactions of experimental group with education, income, index of multiple deprivation, risky drinking, motivation to reduce consumption and self-licensing yielded no statistically conclusive effects (see online Supporting Information Table S3).

Secondary outcome

Self-reported appeal of the wine consumed did not differ conclusively between the three experimental groups; all $Ps > 0.082$ (see online Supporting Information Table S1 for more details).

DISCUSSION

Participants drank significantly more wine when the label contained a numerical descriptor of alcohol strength

Table 3 Regression predicting the rank of total consumption, based on experimental group and stratification covariates ($n_{total} = 147$, $n = 49$ per experimental group).

	Estimate (B)	P value	95% CI
Constant	70.23	<0.0001	52.63 – 87.84
'4% ABV'	20.30	0.016*	3.92 – 36.69
'Super Low 4%ABV'	20.68	0.014*	4.32 – 37.04
(Ref: 'Super Low')			
Gender: female	−19.02	0.006*	−32.42 – −5.62
(Ref: male)			
Age: 45-80 years	−8.74	0.198	−22.11 – 4.63
(Ref: 18–44 years)			
SES: Medium C1C2	7.29	0.378	−9.01 – 23.58
SES: Low DE	5.16	0.538	−11.37 – 21.69
(Ref: High AB)			

*Significant at the 0.0236 threshold (set for the adaptive design).
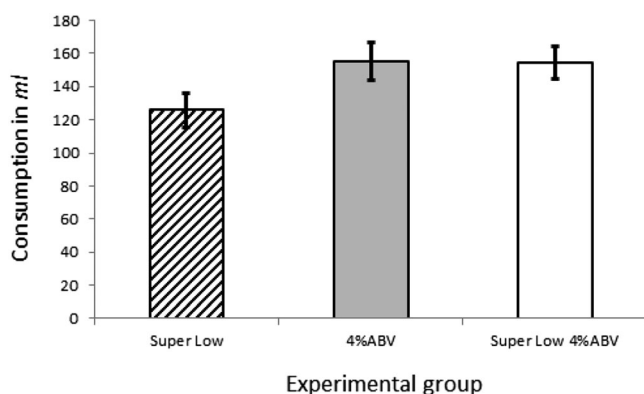
Figure 2 Graphical presentation of consumption levels (ml) across the three experimental groups. Error bars represent ± 1 SE

compared to a label containing only a verbal descriptor of alcohol strength. This higher level of consumption was apparent both when the label contained only a numerical descriptor (4%ABV) and when the label contained a combination of a verbal and numerical descriptor of strength (Super Low 4%ABV).

Taken together, these results suggest that the higher consumption of lower strength alcoholic drinks seen in previous research is driven by the numerical rather than the verbal information of strength on the label. This is in line with recent findings from an online sample of weekly drinkers who relied more heavily on the numerical rather than the verbal information on the label when making judgements as to the target groups and occasions of differently labelled wines and beers [10]. In the principal component analyses (PCA), participants clustered the alcoholic drinks based on the numerical rather than the verbal information on the label. Combined, these findings demonstrate that at least in judgements involving drinks of different alcohol strength numerical information trumps verbal information. Future research should examine the mechanisms that lead consumers to rely more on numerical than verbal information of strength, as well as any

moderating variables to this effect (e.g. individual differences in numeracy levels) [24]).

Although the present study is not a direct replication of the prior bar lab study [11], the experimental group combining verbal and numerical information of lower alcohol strength (Super Low 4%ABV) is identical to the experimental group that displayed the highest level of consumption in the prior study. The level of consumption in both experimental groups containing numerical information of strength in the present study is comparable to the level of consumption for wine in the relevant experimental group in the prior study (in the earlier study consumption was $M = 159.13$ ml, $SD = 84.89$).

Participants in the two experimental groups exposed to numerical information on alcohol strength consumed ~23% more volume of wine when compared to participants not exposed to numerical information of strength. This increased level of consumption may have significance at the population level, although it has to be borne in mind that the present study can only speak to effects arising during a 10-minute taste test in a bar lab setting. Future research is needed to ascertain whether increased

consumption of alcoholic drinks labelled with numerical information of alcohol strength would increase in a linear fashion in naturalistic contexts during prolonged drinking periods.

Participants did not self-report statistically significant differing levels of appeal across the three groups. This speaks to the successful manipulation of our labelling, because participants were all given the same type of wine, with only the label differing between experimental groups. Furthermore, this finding suggests that it is not differences in appeal that lead to increased consumption when labels display lower alcohol strength in numerical format.

The present pattern of results is also in line with prior findings by Geller and colleagues [25] suggesting that labels not highlighting the lower alcohol content of drinks may be more effective in reducing consumption than those in which the lower alcohol content is highlighted. Future research should examine this further.

Strengths and limitations

Although our study shows that presenting information on alcohol strength in numerical format for lower strength wines leads to significantly greater consumption, we do not know whether the same pattern of results would be found for drinks of higher alcohol strength. It is possible that in line with the self-licensing hypothesis [6], numerical information on labels of higher alcohol strength actually leads to lower consumption of alcoholic drinks, because the label may serve as a deterrent cue to moderate the rate of drinking.

The present study entailed a time-limited taste test in a bar laboratory setting. Although the taste test is a commonly used validated measure of alcohol consumption [15], and bar labs provide more ecologically valid testing environments than traditional research lab settings, it would be important to replicate the present findings in field settings (bars and restaurants) to ascertain whether drinkers rely more on the numerical information on the labels in those settings too and what kind of implications this has for patrons' consumption levels. Using naturalistic drinking environments would also enable a measure of what happens when the duration of the consumption episode is not time-limited. It may be that the impact of the numerical information on the label of the low strength alcohol is less pronounced during extended drinking episodes. Replications with other types of alcohol will also be important to ascertain generalisability of the present findings.

POLICY IMPLICATIONS

In policy terms, the present study suggests that policies extending the range of lower strength alcohol labelling may carry unintended consequences, by potentially increasing consumption levels in the population. Although it would be unethical to remove all numerical information on alcohol strength from labels on alcoholic drinks, the present findings coupled with those by Geller and colleagues [25], suggest that policies that encourage the development of lower strength alcohol alternatives, but carrying labels that do not prominently highlight the lower strength of the drinks may be more successful in lowering harmful consumption of alcohol at the population level.

However, it is important to note that the present findings speak only to the potential of overconsumption when drinkers are provided the low alcohol strength drinks. Although recent sales data suggest that the market for low- and no-alcohol drinks is increasing, especially in developed countries, regular/average strength drinks still hold the greatest share of the market [26,27]. Therefore, the potential for lower strength drinks to increase overall consumption at the population level may be counteracted by the lower sales of these products. Furthermore, as prior studies suggest, the paradoxical effects of increased consumption arising from the numerical information provided on the labels of lower strength drinks may in fact work in the opposite direction in purchasing contexts (both on- and off-licence). For example, we know that some segments of the population (e.g. younger and heavier drinkers) use label information to choose higher strength drinks [8]. Further research is needed to disentangle the potential behavioural impact both on purchasing and consumption at the population level arising from any changes to the labelling regulations pertaining to lower strength alcohols.

Trial registration

ISRCTN33451258.

Declaration of interests

The authors declare that they have no competing interests to declare.

Funding

Author contributions

Milica Vasiljevic: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; supervision; visualization. **Daniel Frings:** Conceptualization; investigation; methodology; project administration; supervision. **Mark Pilling:** Conceptualization; data curation; formal analysis; methodology. **Theresa Marteau:** Conceptualization; funding acquisition; investigation; methodology; project administration; resources; supervision.

References

1. WHO. Global status report on alcohol and health 2018. World Health Organization; 2019 Feb 14. Available from https://www.who.int/publications/i/item/9789241565639

2. OECD *Tackling harmful alcohol use*. Paris: OECD Publishing; 2015Available from https://www.oecd.org/health/tackling-harmful-alcohol-use-9789264181069-en.htm

3. Department of Health. Public Health Responsibility Deal–A8 (a) Alcohol Unit Reduction; 2012. Available from https://responsibilitydeal.dh.gov.uk/pledges/pledge/?pl=22

4. Anderson P., Rehm J. Evaluating alcohol industry action to reduce the harmful use of alcohol. *Alcohol Alcohol* 2016 Jul 1; **51**: 383–7; https://doi.org/10.1093/alcalc/agv139

5. Rehm J., Lachenmeier D. W., Llopis E. J., Imtiaz S., Anderson P. Evidence of reducing ethanol content in beverages to reduce harmful use of alcohol. *Lancet Gastroenterol Hepatol* 2016 Sep 1; **1**: 78–83; https://doi.org/10.1016/S2468-1253(16)30013-9

6. Khan U., Dhar R. Licensing effect in consumer choice. *J Market Res* 2006 May; **43**: 259–66; https://doi.org/10.1509/jmkr.43.2.259

7. Bui M., Burton S., Howlett E., Kozup J. What am I drinking? The effects of serving facts information on alcohol beverage containers. *J Consum Aff* 2008 Mar; **42**: 81–99; https://doi.org/10.1111/j.1745-6606.2007.00095.x

8. Jones S. C., Gregory P. The impact of more visible standard drink labelling on youth alcohol consumption: helping young people drink (ir) responsibly? *Drug Alcohol Rev* 2009 May; **28**: 230–4; https://doi.org/10.1111/j.1465-3362.2008.00020.x

9. Vasiljevic M., Coulter L., Petticrew M., Marteau T. M. Marketing messages accompanying online selling of low/er and regular strength wine and beer products in the UK: a content analysis. *BMC Public Health* 2018 Dec; **18**: 1–7; https://doi.org/10.1186/s12889-018-5040-6

10. Vasiljevic M., Couturier D. L., Marteau T. M. What are the perceived target groups and occasions for wines and beers labelled with verbal and numerical descriptors of lower alcohol strength? An experimental study. *BMJ Open* 2019 Jun 1; **9**: e024412; https://doi.org/10.1136/bmjopen-2018-024412

11. Vasiljevic M., Couturier D. L., Frings D., Moss A. C., Albery I. P., Marteau T. M. Impact of lower strength alcohol labeling on consumption: a randomized controlled trial. *Health Psychol* 2018 Jul; **37**: 658–67; https://doi.org/10.1037/hea0000622

12. Chang M. *Adaptive design theory and implementation using SAS and R*. CRC Press; 2014 Dec 1.

13. Lancaster G. A., Dodd S., Williamson P. R. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract* 2004 May; **10**: 307–12; https://doi.org/10.1111/j..2002.384.doc.x

14. Wittes J., Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Stat Med* 1990 Jan; **9**: 65–72; https://doi.org/10.1002/sim.4780090113

15. Jones A., Button E., Rose A. K., Robinson E., Christiansen P., Di Lemma L., *et al.* The ad-libitum alcohol 'taste test': secondary analyses of potential confounds and construct validity. *Psychopharmacology (Berl)* 2016 Mar 1; **233**: 917–24; https://doi.org/10.1007/s00213-015-4171-z

16. Field M., Eastwood B. Experimental manipulation of attentional bias increases the motivation to drink alcohol. *Psychopharmacology (Berl)* 2005 Dec; **183**: 350–7; https://doi.org/10.1007/s00213-005-0202-5

17. 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 Sep 14; **158**: 1789–95; https://doi.org/10.1001/archinte.158.16.1789

18. Public Health England *AUDIT – C, PHE Alcohol Learning Resources*. Public Health England; 2017Available from https://www.alcohollearningcentre.org.uk/Topics/Latest/AU-DIT-C/

19. Oguz S., Merad S., Snape D. *Measuring National Well-being-What matters most to Personal Well-being?*. London: Office for National Statistics; 2013.

20. R Core Team. R *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2017Available from https://www.R-project.org/

21. IBM Corp *IBM SPSS Statistics for Windows, Version 25.0*. IBM Corp: Armonk, NY; 2017.

22. Forsythe A. B. Validity and power of tests when groups have been balanced for prognostic factors. *Comp Stat Data Ana* 1987 Jan 1; **5**: 193–200; https://doi.org/10.1016/0167-9473(87)90015-6

23. Rosenberger W. F., Lachin J. M. *Randomization in clinical trials: theory and practice*. John Wiley & Sons; 2015 Nov 23.

24. Peters E. Beyond comprehension: the role of numeracy in judgments and decisions. *Curr Dir Psychol Sci* 2012 Feb; **21**: 31–5; https://doi.org/10.1177/0963721411429960

25. Geller E. S., Kalsher M. J., Clarke S. W. Beer versus mixed-drink consumption at fraternity parties: a time and place for low-alcohol alternatives. *J Stud Alcohol* 1991 May; **52**: 197–204; https://doi.org/10.15288/jsa.1991.52.197

26. Anderson P., Llopis E. J., O'Donnell A., Manthey J., Rehm J. Impact of low and no alcohol beers on purchases of alcohol: interrupted time series analysis of British household shopping data, 2015–2018. *BMJ Open* 2020 Oct 1; **10**: e036371; https://doi.org/10.1136/bmjopen-2019-036371

27. Woolfson D. Off the wagon: low and no category report 2020. The Grocer; 2020 Apr 14. Available from https://www.thegrocer.co.uk/category-reports/off-the-wagon-low-and-no-category-report-2020/603849.article

Supporting Information

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

Figure S1 Image of the bar lab set-up (participants sat at the bar counter for the intervention period).

Figure S2 Sample label in one experimental group.

Figure S3 Experimental set-up in one of the three experimental groups.

Table S1 Regression modelling the transformed appeal measure (secondary outcome), based on experimental group and stratification covariates (Ntotal = 147, $n = 49$ per experimental group).

Table S2 Regression modelling the rank of total consumption, based on experimental group, age, gender and socio-economic status (SES; occupation) including interactions (Ntotal = 147, $n = 49$ per experimental group).

Table S3 Regression modelling the rank of total consumption, based on experimental group, education, income, index of multiple deprivation, risky drinking, motivation to reduce consumption and self-licensing including interactions (Ntotal = 147, $n = 49$ per experimental group).