



Can incentive-compatibility reduce hypothetical bias in smokers' experimental choice behavior? A randomized discrete choice experiment

John Buckell^{a,*}, Justin S. White^b, Ce Shang^c

^a Health Economics Research Centre, Nuffield Department of Population Health and Nuffield Department of Primary Health Care Sciences, University of Oxford, OX3 7RG, UK

^b Philip R. Lee Institute for Health Policy Studies, University of California, San Francisco, USA

^c Department of Internal Medicine, The Ohio State University Wexner Medical Center, USA

ARTICLE INFO

JEL classification:

C35
I12
I18

Keywords:

Discrete choice experiments
Tobacco
Hypothetical bias
Incentive-compatibility
Willingness to pay

ABSTRACT

Discrete choice experiments (DCEs) are used to provide evidence for policymaking and nonmarket valuation in health. A perennial issue with the stated preference data used in DCEs is hypothetical bias; that is, hypothetical responses in experiments may differ from real-world behavior. A randomized DCE tested whether an incentive-compatible preference elicitation reduced hypothetical bias. Adult smokers were randomly assigned to either an incentive-compatible arm or a control arm; and then made DCE choices among cigarettes, e-cigarettes, and an opt-out. We examined the impacts on product choices, willingness to pay, and the scale of utility. Scale and willingness to pay were unaffected by the incentive. Respondents in the incentive-compatible arm were more likely to choose e-cigarettes. That is, the incentive-compatible approach affected product choices rather than scale/attribute preferences. Thus, while it is feasible to use incentive-compatibility mechanisms to manipulate experimental behaviors, the approach did not induce the hypothesized effect on preferences in this setting.

1. Introduction

Discrete choice experiments (DCEs) are increasingly used in addictive behaviors, particularly in tobacco, to provide evidence for policymaking (Regmi et al., 2018; Soekhai et al., 2018). In tobacco markets, for example, there is urgent need for behavioral evidence in order to inform regulation of cigarettes and e-cigarettes as these markets are changing rapidly (Hu et al., 2019). Typically, researchers collect stated preference (SP) data in these DCEs. A recurrent issue with SP data is hypothetical bias; that is, what individuals say they will do in experimental settings may not correspond with their real-world behaviors (Hensher et al., 2015; Catalogue of Bias Collaboration et al., 2020). This bias can in turn lead to biases in model estimates and thus in the policy evidence derived from these tobacco-based DCEs (Buckell and Hess, 2019).

The effects of hypothetical bias have been studied in the choice modelling literature. In terms of preferences, hypothetical bias has had mixed effects (Fifer et al., 2014). In economic studies, hypothetical bias typically leads individuals to overstate willingness to pay (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005; Whyne et al., 2005; Blumenschein et al., 2008; Donfouet et al.,

* Corresponding author.

E-mail address: john.buckell@ndph.ox.ac.uk (J. Buckell).

2013; Alemu and Olsen, 2018). In the health domain, results from recent studies, including a recent tobacco study, suggest that responses to attribute variation are overstated in the presence of hypothetical bias, i.e., hypothetical bias affects the scale of utility. Furthermore, product choice shares and willingness to pay are vulnerable to hypothetical bias (Lusk and Schroeder, 2004; Özdemir et al., 2009; Buckell and Hess, 2019).

One of the underlying causes of hypothetical bias is when individuals' incentives are misaligned with the researcher's goals and respondents are not motivated to give truthful answers. Responses may be affected by social desirability bias—choosing actions perceived to be desirable—or experimenter demand effects—when this behavior changes with treatment (Zizzo, 2010; de Quidt et al., 2018). Moreover, respondents may simply not feel any binding consequences of their choices. Therefore, we test the impact of an incentive-compatible choice mechanism designed to reduce hypothetical bias in a tobacco-based, randomized DCE.

Incentive-compatibility, a bedrock concept in mechanism design and experimental economics, holds when participants would find it advantageous to reveal their true preferences during a choice task (Hurwicz, 1972; Smith, 1982). Incentive-compatible choice mechanisms have become standard in experiments that measure economic preferences as a way to induce revealed preference behavior (e.g., Holt and Laury, 2002; Andreoni and Sprenger, 2012). The widespread adoption of incentivized choice experiments has been reinforced by a number of studies that have compared choices under hypothetical and real incentives. For example, Holt and Laury (2002) introduced a multiple price list (MPL) design that has become one of the most common approaches for eliciting risk preferences, finding that the number of “safe” choices selected varies for real vs. hypothetical payments. Similarly, Collier and Williams (1999) used an MPL format to test the effects of real vs. hypothetical payments on time preference estimates, finding that discount rates are lower when incentives are real relative to when they are hypothetical. Many if not most economic experiments now use an incentive-compatible mechanism to elicit subjects' revealed preferences (e.g., Andersen et al., 2008; Dohmen et al., 2010). A recent study claims that “incentive-compatible choice experiments ... provide accurate estimates of revealed preferences” (Brynjolfsson et al., 2019), though we note this has not been consistently found in empirical studies (e.g., Lusk and Schroeder, 2004). This approach has not, to our knowledge, been incorporated into DCEs in the health domain. Incentivized experiments, while generally viewed as the gold standard for eliciting preferences, can be expensive and time-consuming. Thus, it is important to understand whether they deliver different estimates from hypothetical elicitations. There has been limited work in applying the concept of incentive-compatibility to the health domain. This is consistent with other recent attempts to understand the generalizability of economic theoretical concepts, such as reference-dependent preferences, ambiguity preferences, and Rabin's paradox to health-related applications (Attema et al., 2013, 2018; Lipman et al., 2019).

In this study, we randomized participants to either a treated, incentive-compatible DCE elicitation, arm or a control, standard DCE elicitation, arm. We then conducted a series of analyses, informed by the prior literature, to study the impact of the incentive-compatible elicitation on product choices, willingness-to-pay estimates, and the scale of utility in the choice model.

2. Methods

2.1. Correcting hypothetical bias in purchasing-based health choices

There are many techniques to correct for hypothetical bias and other forms of bias, e.g. selection bias, in DCEs. The use of these techniques has been limited in health DCEs (Lancsar and Burge, 2014; Buckell and Hess, 2019). These techniques are broadly classified into ex-ante and ex-post techniques (Fifer et al., 2014; Wuepper et al., 2018). Studies in tobacco have applied ex-ante techniques, such as honesty contracts, wherein individuals tick a box to agree to providing honest answers (Buckell and Sindelar, 2019). Elsewhere, several studies take ex-ante approaches in food choices. Wuepper et al. (2018) use both cheap talk and an approach based on respondents' own sense of self-consciousness. Alemu and Olsen (2018) implement two approaches. They first use a revealed preference arm of their experiment to directly elicit truthful responses. Second, they use a repeated opt-out reminder in choice sets, which they go on to show reduces hypothetical bias. Fang et al. (2020) use virtual reality to mitigate hypothetical bias in food choices. They find that hypothetical bias is reduced for those that did not find the virtual reality simulator discomforting.

While some evidence supports the effectiveness of ex-ante techniques, they are not, aside from direct revealed preference approaches, guaranteed solutions for hypothetical bias (Harrison, 2014). Some tobacco studies have applied ex-post techniques, namely econometric calibration using revealed preference data, to correct for hypothetical bias (Kenkel et al., 2020; Buckell and Hess, 2019). Calibration of market shares has also been applied in studies of sugar-sweetened beverages (Blake et al., 2019).

Other available approaches have not yet been tested in purchasing-based DCEs, and may have value to researchers. One of these techniques is to use an incentive-compatible elicitation approach to abate hypothetical bias by conditioning a financial reward on respondents' choices (Ding, 2007; Hensher, 2010; Brynjolfsson et al., 2019). This kind of incentive-compatibility is easy to implement, relatively inexpensive, and can be applied within an experiment, i.e., does not require external data as in the case of calibration, which can be a problem in health (Lancsar and Burge, 2014). Thus, if effective, the approach could be applicable to studies in health that involve purchasing decisions such as tobacco, food, alcohol, illicit substances, and over-the-counter medicines.

2.2. Sample, DCE, and randomization

The experiment was conducted online between August and October 2017. Participants were recruited from Ipsos KnowledgePanel, a national probability online panel. Panel members were recruited using probability selection algorithms for both random-digit dial telephone and address-based sampling methodologies. As such, samples from KnowledgePanel cover all US households regardless of their phone status, and improve population coverage. We recruited 1,154 US adult cigarette smokers who had previously used

electronic cigarettes or vaping products or did not completely rule out future use of these products. The inclusion criteria were intended to avoid respondents who were unlikely to choose e-cigarettes and thus were unlikely to provide any meaningful choice information in the experiment. Table 1 presents descriptive statistics. This sample is broadly similar to past tobacco DCEs (e.g., Marti et al., 2019), which tend to share sample characteristics of all online panels, namely over-representing white and higher income respondents.

In the DCE, respondents chose among labelled options: their usual cigarette product (collected prior to the experiment and piped into the choice sets), two e-cigarette options, and an opt-out. We included “none of the above” because some smokers may have wanted to quit their own cigarettes or switch to alternative tobacco products that are not e-cigarettes, e.g., little cigars or cigarillos. Products were described by five attributes: price, health harm, effectiveness as a cessation aid versus cigarettes, flavor, and nicotine level. These attributes were selected on the basis of prior literature, namely what was known to influence smokers’ decision-making, and what was deemed relevant to policy questions of interest. Using SAS software, a D-optimal design yielded 60 choice sets across five blocks; respondents answered 12 choice tasks each. The experimental design is presented in Table 2, and a sample choice task is given in Fig. 1. Shang et al. (2019) provide full details of the design; and further details are provided in the Supplement.

After some introductory questions in the survey, respondents were randomly assigned in equal proportion to either the treated arm or control arm of the experiment. In the treated arm, respondents were instructed that one of their experimental choices would be selected for implementation among one randomly selected individual. We have provided these instructions in the supplementary materials. For this choice the selected respondent received \$100 of the product chosen. This would have been a dollar reward if respondents chose the opt-out. We leverage this design to test whether the incentives increased disengagement with the experiment, a point that we discuss in the results section. Choices are, therefore, incentive-compatible because utility is maximized by responding truthfully, sometimes termed *random lottery incentives*. Respondents were not instructed of the sample size at the time of randomization, although they were provided with the sample size in the consent form and thus might have had the opportunity to compute the odds of being selected. We did not give the respondents the probability precisely because we wanted the value of the incentives to be more salient than the probability of payoff, thereby strengthening the respondents’ beliefs that it would be better to report accurately and truthfully. In the control arm, no such instructions were given, and no incentive on choices is present.

Our incentive-compatibility approach follows other recent incentivized choice experiments in the tobacco literature, in which a subset of participants receive real payments (Chaloupka et al., 2019). Our design was also influenced by ethical and IRB constraints that steered us away from awarding respondents with free tobacco products. We settled on the current formulation of the choice task to circumvent these constraints.

Baseline characteristics were balanced across arms according to *t*-tests of the differences in means (Table 1).

2.3. Choice modelling

As is common in DCEs, a utility function is defined as capturing individuals’ preferences for tobacco products (Louviere et al., 2000):

$$U_{ijt} = \mu V_{ijt} + \varepsilon_{ijt} \quad (1)$$

$$V_{ijt} = ASC_{cig} + ASC_{ecig} + \beta_p \cdot Price_{ijt} + \beta_h \cdot Heath\ Harm_{ijt} + \beta_c \cdot Cessation_{ijt} + \beta_f \cdot Flavor_{ijt} + \beta_n \cdot Nicotine_{ijt} \quad (2)$$

where U_{ijt} is individual i ’s utility for option j in choice task t . V_{ijt} is the deterministic portion of utility comprising alternative-specific constant terms (ASCs for cigarettes: ASC_{cig} and both e-cigarettes: ASC_{ecig} ; where the opt-out is the reference product), attributes (e.g., price, $Price$) and correspond U_{ijt} ding preferences (β). Attribute levels are dummy-coded (which is tantamount to effects-coding, as used widely in health; see Daly et al., 2016). μ is a scale parameter, inversely proportional to the error term, which governs the influence of the deterministic component of utility on choices (Train, 2009). The scale is confounded with the product/attribute preferences and, as such, is not separable in estimation. For the purposes of estimation, it is assumed to equal 1. ε_{ijt} is the random component of utility assumed to follow an iid type I extreme value distribution (Hensher et al., 2015).

Since we are interested in examining WTP for attributes, we estimate models in the willingness-to-pay space (Train and Weeks, 2005). Note that the models in WTP and preference space are equivalent when the price coefficient is fixed, as is the case here, i.e., the log-likelihoods are identical; thus, the adjustment is superficial. We demonstrate this point in the supplementary materials.

$$V_{ijt} = ASC_{cig} + ASC_{ecig} + \gamma(-Price_{ijt} + \beta_h \cdot Heath\ Harm_{ijt} + \beta_c \cdot Cessation_{ijt} + \beta_f \cdot Flavor_{ijt} + \beta_n \cdot Nicotine_{ijt}) \quad (3)$$

In this form, the attributes are expressed in terms of willingness to pay, though the product ASCs retain their usual interpretation. This model allows us to answer our research questions directly. In addition, we allow for inter-individual heterogeneity by specifying mixing distributions for each of the parameters (Train, 2009). The parameters are then re-specified as a distribution with a mean and a standard deviation to be estimated. Taking health harm as an example,

$$\beta_h = \mu_h + \tau_h \quad (4)$$

Table 1

Descriptive statistics of the sample. Note: The p -values in the last column are taken from chi-squared tests of the association between each categorical variable and randomization group.

	Randomization group			
	Full sample	Control	Treated	p-value
Panel A. Demographics				
Male (%)	50.8	49.7	51.9	0.47
Age (%)				0.96
18-29	13.1	13.4	12.9	
30-44	23.3	22.9	23.7	
45-59	35.8	36.3	35.3	
60+	27.7	27.3	28.1	
Education (%)				0.98
Some high school or less	7.1	7.4	6.8	
High school degree	27.6	27.3	27.8	
Some college	41.5	41.6	41.4	
College degree or more	23.9	23.6	24.1	
Race/ethnicity (%)				0.69
White	75.8	76.7	74.9	
Non-white, non-Hispanic	15.0	14.8	15.3	
Hispanic	9.2	8.5	9.8	
Income (%)				0.58
Less than \$20,000	25.0	24.7	25.3	
\$20,000 - \$39,999	25.4	26.5	24.4	
\$40,000 - \$59,999	15.8	14.6	16.9	
\$60,000 - \$99,999	21.0	20.3	21.7	
\$100,000 or more	12.8	13.9	11.7	
Married (%)	50.1	50.6	49.7	0.75
Urban (%)	84.1	84.7	83.6	0.61
Panel B. Smoking characteristics				
Smokes cigarettes daily (%)	79.1	78.1	80.0	0.43
Vaped ever (%)	77.0	76.9	77.1	0.93
Vapes currently (%)	43.6	43.7	43.4	0.90
No. observations	1 157	567	590	

Table 2

Experimental design.

	Cigarettes	Vape pen
Less harmful to health than cigarettes		Yes [left blank]
Effect for helping people quit		Effective
		Not effective
		Unknown
Nicotine strength	12 mg per stick	None (0 mg)
		Low (1–12 mg)
		Medium (13–17 mg)
		High (18 mg or higher)
Flavor	Respondents' reported flavor	Tobacco
		Menthol
		Fruit/candy/sweet/other flavors
Price	Respondents' reported price per pack	Starter Kit: \$30
		Refill Price:
		\$3
		\$5
		\$7

where μ_h is the mean and $\tau_h \sim N(0, \sigma_{\tau_h}^2)$ is a random component capturing unobserved preference heterogeneity for attribute h . Normal distributions are specified for all attributes, and 2,000 MLHS draws are taken. All models are estimated using Apollo software (Hess and Palma, 2019). We do not estimate a mixing distribution on the price coefficient.¹

To examine if the incentive-compatible elicitation affected product choices or WTP, a variable for whether the individual was exposed to the treatment was interacted with the observed component of utility.

¹ A limitation of this approach is that it imposes a rather strong assumption of the model, namely that the price sensitivity is the same for all the respondents in our sample.

© 2018 GfK - All rights reserved. Need help?

If you were choosing between the following scenarios and these were your only options, which would you choose?

	Cigarettes	Vape Pen A	Vape Pen B
Less harmful to health than cigarettes		Yes	Yes
Effective for helping people quit?		Unknown	Unknown
Nicotine strength	12 mg per stick	High (18 mg or higher)	None (0 mg)
Flavor	Menthol	Tobacco	Tobacco
Price	Price per pack: \$15.00	Starter Kit: \$30 Refill Price: \$3	Starter Kit: \$30 Refill Price: \$3

Select one answer only

☐ Cigarettes
☐ Vape Pen A
☐ Vape Pen B
☐ None of the Above

Fig. 1. Example choice task.

$$V_{ijt} = ASC_{cig} + ASC_{ecig} + \gamma(-Price_{ijt} + \beta_h.Heath\ Harm_{ijt} + \beta_c.Cessation_{ijt} + \beta_f.Flavor_{ijt} + \beta_n.Nicotine_{ijt}) + Incentive_i \times (ASC_{cig,i} + ASC_{ecig,i} + \gamma(-Price_{ijt} + \delta_h.Heath\ Harm_{ijt} + \delta_c.Cessation_{ijt} + \delta_f.Flavor_{ijt} + \delta_n.Nicotine_{ijt})) \quad (5)$$

where $Incentive_i$ is a variable that takes the value of 1 if the individual was exposed to the incentive-compatible elicitation; 0 otherwise. The $ASC_{cig,i}$ and δ_h coefficients capture the marginal effect of treatment on the relative willingness-to-pay/product choice probabilities, and are straightforwardly tested.

To examine the impact of the treatment on the relative scale of utility, a parameter is introduced to estimate differences in scale between treatment and control groups (Hensher and Bradley, 1993; Buckell and Hess, 2019) as below.

$$U_{ijt} = (\mu_{treated}.Incentive_i)(V_{ijt}) + \varepsilon_{ijt} \quad (6)$$

where $Incentive_i$ is as above. $\mu_{treated}$ captures differences in relative scale between the treated and control groups. A test of $\mu_{treated} = 1$ determines any effect of the treatment on the scale (test versus 1 since the scale is assumed to be 1 as noted above).

3. Results

3.1. Main estimates

Estimates of equations (3)–(6) are presented in Table 3. Overall, the estimates suggest that respondents were willing to pay for some of the attributes. Specifically, respondents were willing to pay more than \$3 per pack for increased health—that is, where e-cigarettes are believed to be healthier than cigarettes, rather than unknown (where the level “left blank” is the omitted category). Respondents

Table 3
Mixed logit model estimates in willingness-to-pay space.

	Model 1		Model 2		Model 3	
	Est	Robust t-ratio	Est	Robust t-ratio	Est	Robust t-ratio
E-cigarette 1	2.34	10.57	1.82	6.22	2.27	10.42
E-cigarette 2	1.77	8.16	1.11	3.79	1.65	7.96
Cigarette	6.66	22.10	6.56	18.73	6.70	18.32
s.d.(E-cigarette 1)	-1.43	-9.04	1.37	13.18	1.35	10.50
s.d.(E-cigarette 2)	-0.99	-4.86	0.96	6.96	-0.88	-6.35
s.d.(Cigarette)	4.68	17.62	-4.70	-21.76	-4.75	-16.77
Gamma	0.33	14.31	0.33	8.88	0.31	11.96
E-cigarettes are less harmful to your health	3.43	10.46	3.88	7.27	3.68	9.80
E-cigarettes are effective as a cessation aid: Unknown	1.42	4.97	1.44	3.25	1.47	4.64
E-cigarettes are effective as a cessation aid: Yes	6.57	12.81	6.85	7.97	6.95	11.85
Flavor: Menthol	-3.74	-6.16	-3.20	-3.72	-3.89	-6.76
Flavor: Fruit/candy/sweet/other	-2.50	-5.12	-3.13	-4.36	-2.67	-5.01
Nicotine level: Low	1.24	2.99	1.67	2.84	1.58	3.73
Nicotine level: Medium	0.73	1.67	0.98	1.55	0.84	1.72
Nicotine level: High	-1.72	-3.11	-1.92	-2.36	-2.12	-3.40
s.d.(E-cigarettes are less harmful to your health)	4.42	9.26	-4.89	-8.04	4.99	9.62
s.d.(E-cigarettes are effective as a cessation aid: Unknown)	-1.67	-1.85	-1.82	-3.21	-2.07	-3.23
s.d.(E-cigarettes are effective as a cessation aid: Yes)	-6.13	-9.63	-6.28	-7.70	-6.68	-9.30
s.d.(Flavor: Menthol)	-7.87	-10.16	8.82	8.58	7.79	10.05
s.d.(Flavor: Fruit/candy/sweet/other)	-7.91	-9.33	-8.25	-8.16	-8.72	-10.28
s.d.(Nicotine level: Low)	-3.80	-5.68	-3.57	-4.90	-3.34	-4.37
s.d.(Nicotine level: Medium)	-4.79	-8.47	4.91	7.03	-5.32	-9.58
s.d.(Nicotine level: High)	-6.12	-8.37	6.51	7.13	6.71	9.02
Incentive treatment \times (E-cigarette 1)			1.08	2.19		
Incentive treatment \times (E-cigarette 2)			1.15	2.37		
Incentive treatment \times (Cigarette)			0.67	1.31		
Incentive treatment \times (Gamma)			0.04	0.24		
Incentive treatment \times (E-cigarettes are less harmful to your health)			-0.66	-1.33		
Incentive treatment \times (E-cigarettes are effective as a cessation aid: Unknown)			-0.36	-0.63		
Incentive treatment \times (E-cigarettes are effective as a cessation aid: Yes)			-0.08	-0.11		
Incentive treatment \times (Flavor: Menthol)			-1.24	-1.32		
Incentive treatment \times (Flavor: Fruit/candy/sweet/other)			0.15	0.18		
Incentive treatment \times (Nicotine level: Low)			-0.56	-0.69		
Incentive treatment \times (Nicotine level: Medium)			-0.40	-0.45		
Incentive treatment \times (Nicotine level: High)			0.10	0.10		
mu: Treated arm					1.02	0.41
N	13,753		13,753		13,753	
K	23		35		24	
Log likelihood(fitted)	-8		-8		-8	
	681.25		637.41		650.10	

Note: Standard errors are clustered at the individual level. Model 1 corresponds to Equation (3); Model 2 corresponds to Equation (5); and Model 3 corresponds to Equation (6). T-ratios are for 0 except for in the case of the scale parameter, which is for 1.

Table 4
Predicted choice shares from Model 2.

	Control	Treatment	Difference
Cigarette	0.605	0.593	-2.0%
E-cigarette	0.301	0.342	13.6%
Opt-out	0.094	0.065	-30.9%

Choice shares of products from model 2 (shown as it is the model that allows for differences in choice shares). The difference reflects the effect of the incentive. Therefore, we show the difference in choice share from the control compared to the treatment.

were willing to pay more than \$6 per pack if e-cigarettes are effective as a cessation aid, and slightly less, around \$1 per pack, for low/medium levels of nicotine. Respondents were willing to accept non-tobacco flavors if compensated by around \$4 per pack; and high nicotine if compensated by around \$2 per pack. The significant mixing distributions suggest that willingness to pay and willingness to accept vary around the mean of the estimated means over individuals. From the ASCs, respondents preferred their usual cigarettes and e-cigarettes to the opt-out option; and these preferences vary across individuals, as reflected in the statistically significant standard deviations of the mixing distributions.

In our main interaction models (Model 2), we interact the incentive-compatible treatment with the utility function. Overall, we do not observe any effect of the treatment on respondents' willingness to pay for attributes, as evidenced by the interaction terms. However, the e-cigarette product interactions were significant at the 5% level for both e-cigarette options; e-cigarettes were chosen more often when respondents were incentivized. In simulations, e-cigarettes gained around 13% of its choice share in the treated arm compared to the control arm, see Table 4. We also note in passing that treated individuals still preferred their own cigarette product overall; it is the relative preferences that were affected.

In mixed logit with scale parameter model (Model 3), we test the impact of the incentive-compatible treatment on the scale of utility, i.e., the relative scale of the experimental arm versus the control arm. Thus, we are testing if the parameter $\mu_{treated}$ is statistically different from 1, the assumed scale of the control arm. We fail to reject the null hypothesis ($H_0 : \mu_{treated} = 1$) at any acceptable level of statistical significance. We conclude that the incentive-compatible elicitation did not change the scale of utility in the choice experiment.

3.2. Robustness and sensitivity

Several steps were taken to examine the robustness of our findings. We separately tested the treatment interactions of the ASCs and attributes. We allowed for observed heterogeneity by interacting the cigarette and e-cigarette ASCs with sets of sociodemographic characteristics: age, gender, ethnicity and education. The main results listed above held. In preliminary testing, we found statistically significant differences between preferences for the two e-cigarette options. We note that in such circumstances, this can reflect left-to-right bias in choice scenarios (Hensher et al., 2015). This could have been the case here since the left option had a larger coefficient across all models. It does not appear to have affected the incentive, however: the interactions are both significant and of similar magnitude (we did not reject equality in testing of the coefficients). For this reason we retained the specification in Table 3. We have chosen to present the simpler mixed logit models for ease of exposition, and the results of more highly specified models were not materially different from those presented here. Because we used respondents' own cigarette price in the choice sets, we can use this information in the model to infer whether respondents compared the prices in the experiment to the price of their usual cigarette. For this, we used a model with reference-dependent price sensitivity (Holte et al., 2016). The specification and results of this testing are shown in Supplement D. This specification failed to improve the model fit relative to the basic mixed logit model.

Embedded in the design is an additional feature that allows us to test individuals' behavior in response to the incentive. Specifically, the opt-out option, "none of the above" was also incentivized. Thus, if respondents chose the opt-out in all choice sets, they would increase their chance to receive \$100 cash. This allows us to observe if the incentive induced disengagement with the experiment in favor of receiving a cash reward. In the treated arm, of the 590 respondents, we find that only 5 chose the opt-out option in all choice sets. In the unincentivized control arm, 10 individuals chose the opt-out option in all choice sets. Therefore, we do not believe that the incentive induced disengagement with the experiment.

4. Discussion

We tested the impact of an incentive-compatible elicitation on smokers' experimental behavior to reduce hypothetical bias. We found that the incentive-compatible elicitation encouraged the selection of e-cigarettes; but did not impact on willingness to pay for attributes, nor on the scale of utility.

Prior literature suggests that willingness to pay is often overstated in experimental settings (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005; Whynes et al., 2005; Donfouet et al., 2013). Accordingly, if incentive-compatibility reduced hypothetical bias, we should expect to see a reduction in willingness to pay in the treated arm. In this setting, we found no evidence of a reduction in willingness to pay in the treatment arm.

Next, the literature suggests that, in general, hypothetical bias leads to exaggerated responses to attribute variation in choices (Train, 2009; Hensher et al., 2015; Buckell and Hess, 2019). Accordingly, if incentive-compatibility reduced hypothetical bias, we should expect to see a reduction in scale in the treated arm. There was no variation in scale found between arms. These findings could have several explanations. First, it would be consistent with our findings if, aside from reducing hypothetical bias, the incentive-compatible elicitation induced some other behavior, namely choosing e-cigarettes more often. Second, it is possible that the control group did not suffer from hypothetical bias and so there was limited scope for the incentive-compatible elicitation to alter choice behavior. This is in keeping with three ideas: first, that the design pivoted around the respondents' own products embeds familiarity for respondents and is thought to reduce hypothetical bias (McFadden, 2014; Hensher et al., 2015 ch.19); second, that respondents either use, or intend to use, e-cigarettes so the task was perceived as realistic, thus mitigating hypothetical bias (Hensher et al., 2015 ch.19); and third, that respondents were all instructed that their answers were being used for research, which may have induced realistic responses. Such "cheap talk" is sometimes effective in DCEs, but typically only among inexperienced consumers (List and Gallet, 2001; Lusk, 2003; Carlsson et al., 2005). However, that the experiment did not suffer at all from hypothetical bias seems

unlikely, and its absence cannot be said for certain. Ultimately, we have tested incentivized stated preferences vs. unincentivized stated preferences, which may not see behaviors shift in the expected revealed preference direction(s).

Our results indicate that e-cigarettes were chosen more frequently in the incentivized arm. This finding seems plausible for several reasons related to our sampling of smokers who did not reject e-cigarettes. First, respondents who do not regularly use e-cigarettes may have viewed the incentive offer as a way to get a “free trial” of e-cigarettes. The kit price for e-cigarettes can be expensive and may be a barrier to substitution for e-cigarettes among some smokers (Huang et al., 2014; Cuomo, 2016; Marti et al., 2019). Second, respondents who do regularly use e-cigarettes may have viewed the incentive offer as an opportunity to get e-cigarettes for free. This could either reflect that the perceived price for e-cigarettes is more expensive than for cigarettes (Thirlway, 2016), or perhaps a mental accounting effect, whereby smokers consider the potential \$100 of e-cigarettes as a windfall in a smoking account (Thaler, 1999).

A final issue, perhaps a limitation, is the strength of our incentive-compatibility treatment. In this study, the incentive-compatibility mechanism, while providing incentives for accurate reports, is fairly weak, due to the fact that only a single respondent was eventually given the reward. Though the chances for a given individual receiving the reward are small, we note research showing that respondents' incentives appear to be maintained even when the chances of getting the reward decline (Harrison and Rutstrom, 2008). Also note that while respondents were told the sample size in the consent form, we suspect that the information was not salient to them later during the experiment, and so whether they used this information to calculate their odds of winning. (Of course, if they did, they would have underestimated the odds since only half of the sample was allocated to the treated group.) Also note that some response to the incentive was observed. If the treatment had no effect, then incentives may have been an issue, but since an effect was observed, it seems less likely that the strength of the incentive was the problem. Further, this general class of incentive mechanism is commonly used and has been validated in other contexts (Cubitt et al., 1998); though random lottery experiments are not perfect (Starmer et al., 1991; Cox et al., 2015). In addition, studies have found weak incentive-compatibility to perform well for tobacco purchase tasks (Wilson et al., 2015; Chaloupka et al., 2019). The effect size might have increased if more respondents were given the reward or a larger reward was offered. Whether more potent incentives would induce similar behaviors remains unknown.

In this study, we show that incentive-compatible elicitation can feasibly be used in DCEs with consumer demand type health settings to induce behavioral change when measuring preferences. We do not, however, find evidence to suggest that the incentive mitigates the potential hypothetical bias. However, some plausible responses to the incentive-compatible elicitation, in light of the sample and experimental features, were observed. Whether these results are generalizable to other health settings is unclear. Therefore, we conclude that in this setting, the hypothesized effect does not appear to have been induced. As discussed above, altering the design of the incentive-compatibility mechanism may be one way to alter behaviors in the theoretically expected way. Had we used a different incentive design, different effects and conclusions might have been reached. Future research could use multiple experimental arms, each varying the incentive design, to shed light on this issue.

In consumer demand type health settings, another technique to mitigate hypothetical bias in tobacco DCEs and other experiments is to combine both stated preference (SP) and revealed preference (RP) data to calibrate results (see e.g. Buckell and Hess, 2019; Blake et al., 2019). Hypothetical bias could be mitigated by anchoring SP market share estimates to reflect real-world market shares; and/or rescaling the estimates of attributes that can be estimated using both SP and RP data so that their scales align (e.g., price elasticities). However, this technique depends on the availability of valid RP data, which could be a challenge because of the lack of vape shop sales data and a consensus on the price elasticity of e-cigarette demand. In other words, just because this data is available, doesn't guarantee that it will fix the hypothetical bias problem (e.g. if SP and RP data are not well-suited; see Buckell and Hess, 2019). Therefore, researchers may want to weigh pros and cons when applying techniques to mitigate hypothetical bias. Other approaches include embedding RP responses in the experiment, such as Alemu et al. (2017). Of course, the natural drawback here is that this is a resource intensive approach that may not be feasible. Moreover, if the RP element of the design is still different to real life choice making, then the approach may not entirely eradicate the issue. Other ex-ante approaches, such as honesty priming or cheap talk are also available. As would be certainty calibration. But again, whether these approaches solve the issue of hypothetical bias is not clear. A promising approach might be to use virtual reality, as in Fang et al. (2020). This seems to be a fruitful area for future research.

CRediT authorship contribution statement

John Buckell: Methodology, Software, Formal analysis, Writing - original draft, Writing - review & editing. **Justin S. White:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition. **Ce Shang:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition.

Declaration of competing interest

None.

Acknowledgements

The project is funded by grant number P50DA036128 from the NIH/NIDA and FDA Center for Tobacco Products (CTP). Additional funding came from the state of Oklahoma under grant, “Discrete choice experiments in tobacco regulatory science. Oklahoma Tobacco Settlement Endowment Trust (TSET). Oklahoma Tobacco Research Center, University of Oklahoma Health Sciences.” The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, TSET, or the Food and Drug

Administration. This study was not pre-registered. As such, results should be considered exploratory. CS was also supported by the NIAAA grant 1K99AA024810. The authors thank Scott Weaver, James Nonnemaker, and Jidong Huang for the initial development of this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jocm.2020.100255>.

References

- Alemu, M.H., Olsen, S.B., 2018. Can a Repeated Opt-Out Reminder mitigate hypothetical bias in discrete choice experiments? An application to consumer valuation of novel food products. *Eur. Rev. Agric. Econ.* 45 (5), 749–782.
- Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E.E., 2008. Eliciting risk and time preferences. *Econometrica* 76 (3), 583–618.
- Andreoni, J., Sprenger, C., 2012. Estimating time preferences from convex budgets. *Am. Econ. Rev.* 102 (7), 3333–3356.
- Attema, A.E., Brouwer, W.B.F., l'Haridon, O., 2013. Prospect theory in the health domain: a quantitative assessment. *J. Health Econ.* 32 (6), 1057–1065.
- Attema, A.E., Bleichrodt, H., l'Haridon, O., 2018. Ambiguity preferences for health. *Health Econ.* 27 (11), 1699–1716.
- Blake, M.R., Lancsar, E., Peeters, A., Backholer, K., 2019. Sugar-sweetened beverage price elasticities in a hypothetical convenience store. *Soc. Sci. Med.* 225, 98–107. <https://doi.org/10.1016/j.socscimed.2019.02.021>.
- Blumenschein, K., et al., 2008. Eliciting willingness to pay without bias: evidence from a field experiment*. *Econ. J.* 118 (525), 114–137.
- Brynjolfsson, E., Collis, A., Eggers, F., 2019. Using massive online choice experiments to measure changes in well-being. *Proc. Natl. Acad. Sci. Unit. States Am.* 116 (15), 7250. <https://doi.org/10.1073/pnas.1815663116>.
- Buckell, J., Hess, S., 2019. Stubbing out hypothetical bias: improving tobacco market predictions by combining stated and revealed preference data. *J. Health Econ.* 65, 93–102.
- Buckell, J., Sindelar, J.L., 2019. The impact of flavors, health risks, secondhand smoke and prices on young adults' cigarette and e-cigarette choices: a discrete choice experiment. *Addiction*. <https://doi.org/10.1111/add.14610>.
- Carlsson, F., Frykblom, P., Johan Lagerkvist, C., 2005. Using cheap talk as a test of validity in choice experiments. *Econ. Lett.* 89 (2), 147–152.
- Catalogue of Bias Collaboration, Buckell, J., Buchanan, J., Wordsworth, S., Becker, F., Morrell, L., Roope, L., Kaur, A., Abel, L., 2020. Hypothetical Bias. In: Catalogue of Bias. <https://catalogofbias.org/biases/hypothetical-bias/>. (Accessed 25 October 2020).
- Chaloupka, I.V., Frank, J., Levy, Matthew R., White, Justin S., 2019. Estimating Biases in Smoking Cessation: Evidence from a Field Experiment. No. w26522. National Bureau of Economic Research. <https://www.nber.org/papers/w26522>.
- Cox, James C., Sadiraj, Vjollca, Schmidt, Ulrich, 2015. "Paradoxes and mechanisms for choice under risk. *Exp. Econ.* 18 (2), 215–250.
- Cubitt, Robin P., Starmer, Chris, Sugden, Robert, 1998. On the validity of the random lottery incentive system. *Exp. Econ.* 1 (2), 115–131.
- Cuomo, R.E., Miner, A., Mackey, T.K., 2016. Pricing and sales tax collection policies for e-cigarette starter kits and disposable products sold online. *Drug Alcohol Rev.* 35 (1), 110–114.
- Daly, A., Dekker, T., Hess, S., 2016. Dummy coding vs effects coding for categorical variables: clarifications and extensions. *Journal of Choice Modelling* 21, 36–41.
- de Quidt, J., Haushofer, J., Roth, C., 2018. Measuring and bounding experimenter demand. *Am. Econ. Rev.* 108 (11), 3266–3302.
- Ding, M., 2007. An incentive-aligned mechanism for conjoint analysis. *J. Market. Res.* 44 (2), 214–223.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2010. Are risk aversion and impatience related to cognitive ability? *Am. Econ. Rev.* 100 (3), 1238–1260.
- Donfouet, H.P.P., Mahieu, P.-A., Malin, E., 2013. Using respondents' uncertainty scores to mitigate hypothetical bias in community-based health insurance studies. *Eur. J. Health Econ.* 14 (2), 277–285.
- Fang, D., Nayga Jr., R.M., West, G.H., Bazzani, C., Yang, W., Lok, B.C., Snell, H.A., 2020. On the use of virtual reality in mitigating hypothetical bias in choice experiments. *Am. J. Agric. Econ.* <https://doi.org/10.1111/ajae.12118>.
- Fifer, S., Rose, J., Greaves, S., 2014. Hypothetical bias in Stated Choice Experiments: is it a problem? And if so, how do we deal with it? *Transport. Res. Pol. Pract.* 61, 164–177.
- Harrison, G., 2014. Real choices and hypothetical choices. In: Hess, S., Daly, A. (Eds.), *Handbook of Choice Modelling*. Edward Elgar Publishing, Cheltenham.
- Harrison, G., Elisabet Rutström, E., 2008. Risk aversion in the laboratory. In: Cox, J., Harrison, G. (Eds.), *Risk Aversion In Experiments (Research In Experimental Economics)*, vol. 12. Emerald Group Publishing Limited, Bingley, pp. 41–196.
- Hensher, D.A., 2010. Hypothetical bias, choice experiments and willingness to pay. *Transp. Res. Part B Methodol.* 44 (6), 735–752.
- Hensher, D.A., Bradley, M., 1993. Using stated response choice data to enrich revealed preference discrete choice models. *Market. Lett.* 4 (2), 139–151.
- Hensher, D., Rose, J.M., Greene, W., 2015. *Applied Choice Analysis*, second ed. Cambridge University Press, Cambridge.
- Hess, S., Palma, D., 2019. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling* 32, 100170.
- Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. *Am. Econ. Rev.* 92 (5), 1644–1655.
- Holte, J.H., Sivey, P., Abelsen, B., Olsen, J.A., 2016. Modelling nonlinearities and reference dependence in general practitioners' income preferences. *Health Econ.* 25 (8), 1020–1038. <https://doi.org/10.1002/hec.3208>.
- Hu, S.S., Homa, D.M., Wang, T., Gomez, Y., Walton, K., Lu, H., Neff, L., 2019. State-specific patterns of cigarette smoking, smokeless tobacco use, and E-cigarette use among adults—United States, 2016. *Prev. Chronic Dis.* 16.
- Huang, J., Tauras, J., Chaloupka, F.J., 2014. The impact of price and tobacco control policies on the demand for electronic nicotine delivery systems. *Tobac. Contr.* 23 (3), iii41.
- Hurwicz, Leonid, 1972. On informationally decentralized systems. In: McGuire, C.B., Radner, R. (Eds.), *Decision and Organization: A Volume in Honor of Jacob Marschak*. North Holland, Amsterdam.
- Kenkel, D.S., Peng, S., Pesko, M. F., Wang, H., 2020. Mostly harmless regulation? Electronic cigarettes, public policy, and consumer welfare. *Health Econ.* 29 (11), 1364–1377.
- Lancsar, E., Burge, P., 2014. Choice modelling research in health economics. In: Hess, S., Daly, A. (Eds.), *Handbook of Choice Modelling*. Edward Elgar Publishing, Cheltenham.
- Lipman, S.A., Attema, A.E., 2019. Rabin's paradox for health outcomes. *Health Econ.* 28 (8), 1064–1071.
- List, J.A., Gallet, C.A., 2001. What experimental protocol influence disparities between actual and hypothetical stated values? *Environ. Resour. Econ.* 20 (3), 241–254.
- Little, J., Berrens, R., 2004. Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis. *Econ. Bull.* 3 (6), 1–13.
- Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. *Stated Choice Methods: Analysis and Applications*/Jordan J. Louviere, David A. Hensher, Joffre Swait, with a Contribution by Wiktor Adamowicz. Cambridge, UK; New York, NY, US.
- Lusk, J.L., 2003. Effects of cheap talk on consumer willingness-to-pay for golden rice. *Am. J. Agric. Econ.* 85 (4), 840–856.
- Lusk, J.L., Schroeder, T.C., 2004. Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *Am. J. Agric. Econ.* 86 (2), 467–482.
- Marti, J., Buckell, J., Maclean, J.C., Sindelar, J., 2019. TO "vape" or smoke? Experimental evidence ON adult smokers. *Econ. Inq.* 57 (1), 705–725.

- McFadden, D., 2014. The new science of pleasure: consumer choice behavior and the measurement of well-being. In: Hess, S., Daly, A. (Eds.), *Handbook of Choice Modelling*. Edward Elgar, Cheltenham.
- Murphy, J.J., et al., 2005. A meta-analysis of hypothetical bias in stated preference valuation. *Environ. Resour. Econ.* 30 (3), 313–325.
- Özdemir, S., Johnson, F.R., Hauber, A.B., 2009. Hypothetical bias, cheap talk, and stated willingness to pay for health care. *J. Health Econ.* 28 (4), 894–901.
- Regmi, K., Kaphle, D., Timilsina, S., Tuha, N.A.A., 2018. Application of discrete-choice experiment methods in tobacco control: a systematic Review. *Pharmacoeconomics - Open* 2 (1), 5–17.
- Shang, C., Weaver, S., Huang, J., White, J.N.F.C.J., 2018. Product Preference Among Adult Smokers – A Discrete Choice Experiment.
- Smith, Vernon L., 1982. Microeconomic systems as an experimental science. *Am. Econ. Rev.* 72 (5), 923–955.
- Soekhai, V., de Bekker-Grob, E.W., Ellis, A.R., Vass, C.M., 2018. Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics*. <https://doi.org/10.1007/s40273-018-0734-2>.
- Starmer, Chris, Sugden, Robert, 1991. Does the random-lottery incentive system elicit true preferences? An experimental investigation. *Am. Econ. Rev.* 81 (4), 971–978.
- Thaler, R.H., 1999. Mental accounting matters. *J. Behav. Decis. Making* 12 (3), 183–206.
- Thirlway, F., 2016. Everyday tactics in local moral worlds: E-cigarette practices in a working-class area of the UK. *Soc. Sci. Med.* 170, 106–113.
- Train, K., 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.
- Train, K., Weeks, M., 2005. Discrete Choice Models in Preference Space and Willingness-To-Pay Space. *Applications of Simulation Methods in Environmental and Resource Economics*. In: Scarpa, R., Alberini, A. (Eds.). Springer Netherlands, Dordrecht, pp. 1–16.
- Whynes, D.K., et al., 2005. Think of a number any number? *Health Econ.* 14 (11), 1191–1195.
- Wilson, A. George, et al., 2015. Behavioral economics of cigarette purchase tasks: within-subject comparison of real, potentially real, and hypothetical cigarettes. *Nicotine Tob. Res.* 18 (5), 524–530.
- Wuepper, D., Clemm, A., Wree, P., 2019. The preference for sustainable coffee and a new approach for dealing with hypothetical bias. *J. Econ. Behav. Organ.* 158, 475–486.
- Zizzo, Daniel John, 2010. Experimenter demand effects in economic experiments. *Exp. Econ.* 13 (1), 75–98.