Online Appendix: Fear or Knowledge? The Impact of Graphic Cigarette Warning on Tobacco Product Choices

In this Appendix we first check the robustness of the main results reported in the paper, and then discuss several lines of evidence that shed light on the validity of the stated preference (SP) data we collected through our discrete choice experiment (DCE), and the implications for the empirical results reported in the text of the paper.

In section A we estimate a multinomial logit model of the probabilities of cigarettes, e-cigarettes, and quitting choices, and calculate the average marginal effects as a robustness check for the main results estimated from linear probability models reported in the paper. In section B we review previous research that compares SP and revealed preference (RP) data in a range of applications. In section C we review previous research that conducts DCEs of tobacco product choices. In Section D we provide additional empirical evidence on the validity of our SP data. Based on previous research and the empirical evidence in section D, we conclude that because tobacco products are familiar market goods, econometric models estimated using the SP data from our DCE are likely to provide reliable forecasts of consumer demand.

A. Robustness of Main Results

Table A1 provides estimated results from a multinomial logit model of the probabilities of cigarette, e-cigarette, and quitting choices. The marginal effects of product attributes are almost identical to the linear probability model estimates presented in Table 2.

Table A1. Marginal Effects Derived from a Multinomial Logit Model

Variables		Immediate choice today			Choice of 6 months from now		
		Cigarette	E-cigarette	Quit	Cigarette	E-cigarette	Quit
Cigarette warning	GWL amputation	-0.053***	0.028*	0.025	-0.076***	0.024	0.052**
		(0.018)	(0.016)	(0.016)	(0.019)	(0.017)	(0.021)
Price	Cigarette price	-0.018***	0.010***	0.008***	-0.016***	0.008***	0.009***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	E-cigarette price	0.017***	-0.017***	0.000	0.012***	-0.014***	0.002*
		(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
E-cigarette available flavor	Tobacco, menthol, fruit/sweet/candy	-0.023***	0.028***	-0.005	-0.027***	0.036***	-0.009
		(0.008)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)
	Tobacco and menthol	-0.025***	0.010	0.016**	-0.021***	0.019***	0.003
		(0.008)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)
E-cigarette available nicotine level	Up to 20mg	-0.004	0.001	0.003	-0.004	-0.003	0.006
		(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)
	Up to 50mg	0.011*	-0.010	-0.001	-0.002	-0.001	0.003
		(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
E-cigarette warning	Are not completely risk free	0.005	-0.002	-0.003	0.001	0.002	-0.003
		(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)
	Contain nicotine, which is addictive	0.032***	-0.017**	-0.014*	0.015*	-0.004	-0.011
		(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
	May expose users to chemicals and toxins	0.012	-0.016**	0.004	0.006	-0.011	0.005
		(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)

Observations	14,052	14,052
Log-likelihood at convergence	-11413.185	-12372.035

Notes: Additional control variables include smoking status (being every day smoker), age of smoking initiation, heavy smokers (heavy smoker index >= 3), vaping status (ever used e-cigarette), quitting effort (tried to quit during last 12 months), quit intention (intend to quit in the next 6 months), gender (being male), age, household size, education (reference category is grade school/some high school), household income above 50k, full-time employed, white. The reference category of e-cigarette available flavor "tobacco only", the reference category of e-cigarette available nicotine level is "up to 5mg", the reference category of e-cigarette warning is "no warning". Standard errors are calculated by the Delta method. *** p<0.01, ** p<0.05, * p<0.1.

B. Literature Review of Studies that Compare SP and RP Data

DCEs and the related contingent valuation method are used to collect SP data in a range of applications. DCEs are commonly used in marketing research and economics to provide predictions of consumer demand in scenarios that are not yet observed in actual markets. In addition to the tobacco product DCEs discussed below in section C, examples of the use of DCEs to study hypothetical market situations include studies of electricity markets (Blass, Lach, and Manski, 2010), health insurance markets (Kesternich, Hiess, McFadden, and Winter, 2013), labor markets (Mas, Alexandre and Pallais, 2017; Maestas et al. 2023), and firearms markets (Moshary, Shapiro, & Drango NBER Working Paper 2023). DCEs are also widely used in health economics to evaluate existing or prospective pharmaceutical products and health care treatment interventions (Ryan et al. 2007). Another large body of research uses DCEs and the related contingent valuation method to estimate willingness to pay for non-market goods like environmental quality.

Research that compares SP and RP data concludes that the external validity of SP data is much stronger in applications similar to familiar market goods. In a narrative review of DCE research, McFadden (2017) concludes that there is a "sharp reliability gradient":

> Forecasts that are comparable in accuracy to RP forecasts can be obtained from well-designed SP studies for familiar, relatively simple goods that are similar to market goods purchased by consumers, particularly when calibration to market benchmarks can be used to correct experimental distortions. However, studies of unfamiliar, complex goods give erratic, unreliable forecasts.

McFadden is therefore skeptical about SP data on complex and unfamiliar environmental public goods. Although he does not discuss health care applications, by the same reasoning DCEs might not provide reliable data on unfamiliar pharmaceutical and health care treatment interventions.

Penn and Hu (2018) report a meta-analysis that provides quantitative evidence consistent with McFadden's (2017) conclusion that SP data are more reliable for familiar market goods. The meta-analysis used estimates from 132 studies that provided 908 observations of comparisons of SP and RP data. For studies including choice experiments that did not provide estimates of willingness to pay but did provide proportions of responses, Penn and Hu inferred lower-bound estimates of willingness to pay. Each observation is an estimate of the "calibration factor" (CF) which shows the ratio of willingness to pay estimated from SP data to the willingness to pay estimated from RP data. When SP and RP estimates are similar, the CF will be close to one. In the meta-analysis, about one quarter of the CFs are between 0.81 and 1.2. The distribution of CFs is skewed right showing a tendency for SP willingness-to-pay estimates to be larger than RP estimates, sometimes to a large extent. The median CF is 1.94, implying that for almost half of the observations the SP estimate is over twice as large as the RP estimate. Penn and Hu estimate regression models of the effects of study characteristics on CFs. The results imply that compared to studies of public goods, studies of private goods find lower CFs. The results also imply that compared to other hypothetical elicitation methods, CFs are lower for studies that used choice experiments. The meta-analysis empirical results are consistent with McFadden's conclusion that although SP data from DCEs are reliable for private goods, there is a sharp reliability gradient for SP data on willingness to pay for public goods.

In the remainder of this section, we discuss examples of studies that compare SP and RP data on health-related choices. Quaife et al (2018) review a number of studies of health-related choices that focused on the external validity of DCEs (by comparing SP and RP estimates) and conclude that DCEs provide moderate levels of external validity in terms of matching actual choices. de Bekker-Grob et al (2020) find that when measured at the individual level, stated preferences in a DCE about vaccinations predict 91 percent of actual choices. Telser and Zweifel (2007) examine the external validity of a DCE focused on decisions about a harm reduction product (hip protectors for accidental falls). They compare the willingness to pay for risk reduction that was derived from the DCE to other measures of willingness to pay for the same risk reduction derived from established alternatives that used revealed preference data. The comparison supports a high level of convergent validity. Linley and Hughes (2013) examine hypothetical decisions about new medicine approvals and find that the predicted probabilities of recommending new medicines derived from the DCE match well with the cumulative probability of actual positive recommendations (though the ability of the DCE to discriminate between individual new medicines was limited). Mahammad et al (2017) use a DCE and estimate a mixed logit based on the hypothetical choices with respect to type of tuberculosis treatment (or none) in response to each treatment having six treatment attributes. They compare these choices with actual choices and find strong external validity and the degree of accuracy depends on the distributional assumptions used in the mixed logit models with some models. Kesternich et al (2013) implement a DCE to analyze Medicare part D choices and compare these results to those that emerge from analysis of actual choices. They conclude that hypothetical choice experiments are useful in studying insurance choices as hypothetical behavior is related to actual behavior. They find that the coefficients that emerge in the DCE experiment are of the same sign as the coefficients that are estimated from market behavior. They note that the magnitudes of the coefficients are quite similar and do not find significant differences between hypothetical and real choices between different attributes of the insurance scenarios. They do find a higher willingness to pay for insurance in the hypothetical market and thus higher insurance take-up rates but attribute this to the nature of the default option in the DCE.

C. DCE Studies of Tobacco Product Choices

A growing body of research conducts DCEs to study the determinants of consumer choices about tobacco products. Table C1 lists recent DCE studies of tobacco product choices published in economics journals. None of these studies compare graphic warning labels to text warning labels; however, we discuss below an economics study of graphic warning labels that used the results of non-hypothetical experimental auctions. We also discuss three studies published in public health journals that use methods similar to DCEs to estimate the impact of different sizes of graphic warning labels.

Table C1 includes two studies that report results from exercises that use a combination of RP and SP data to develop a calibrated model that is grounded in real-world behavior. Kenkel, Peng, Pesko, and Wang (2020) report in an Online Appendix the results of a conditional logit model estimated using a combination of SP data from a DCE and RP data from the DCE subjects' responses about their prior use of e-cigarettes, combustible cigarettes, and nicotine replacement products. The estimated scale parameter is close to 1, suggesting that the scales in the RP and SP data are similar. The estimated coefficients on the tobacco product attributes show the same patterns as in the model based on SP data only.

Buckell and Hess (2019) report the results of a more in-depth investigation of combining SP and RP data on tobacco product choices. In a model estimated using combined SP and RP data they estimate a scale parameter greater than 1, consistent with the argument that in SP data subjects overstate the impact of interventions which leads to elasticities that are biased upwards. In terms of forecasts from the models, they find that compared to calibrated models the uncalibrated forecasts under-predict cigarette choices and over-predict e-cigarette choices. They conclude that appropriately calibrated choice models "provide better quality empirical evidence for policymakers." (Buckell and Hess 2019, p. 100)

The remaining tobacco product DCEs listed in Table C1 do not provide in-depth discussions of external validity, but the results of the studies are consistent with predictions from health economic models of consumer behavior. For example, Marti, Buckell, Maclean, and Sindelar (2019) conduct a DCE to study how smokers' product choices are affected by variations in the perceived healthiness and cessation effectiveness of e-cigarettes, as well as by bans on smoking in public places and prices. Buckell, Hensher, and Hess (2021) use SP data from a DCE combined with a latent variable approach to model addiction. They find that more addicted smokers are unwilling to switch to e-cigarettes. Guindon, Metnzakis, and Buckley (2024) conduct a DCE to study the impact of plain cigarette packaging and text warnings on cigarette sticks. The study is set in Canada, where graphic warning labels were already required. Almost half the subjects always chose cheaper illicit cigarettes that did not carry the required graphic warning; however, it is difficult to disentangle the price and label effects. The warnings on cigarette sticks discouraged product choices but did not significantly influence health risk perception. More broadly, the studies in Table C1 find that smokers' product choices respond to cigarette and e-cigarette prices; the estimated price-responsiveness is generally consistent with the large body of econometric estimates from observational (RP) data (DeCicca, Kenkel, and Lovenheim 2022).

Table C1. Recent DCE Studies of Tobacco Product Choices in Economics Journals

Authors	Year	Journal
Buckell, Hensher, and Hess	2021	Health Economics
Buckell and Hess	2019	Journal of Health Economics
Guindon, Mentzakis, and Buckley	2024	Economics & Human Biology
Kenkel, Peng, Pesko, and Wang	2020	Health Economics
Marti, Buckell, Maclean, and Sindelar	2019	Economic Inquiry

In addition to the tobacco product DCEs listed in Table C1, Rousu, Marette, Thrasher, and Lusk (2014) report the results of non-hypothetical experimental auctions where smokers placed bids on cigarette packages that carried either a text-only or graphic warning label. Depending on the location of the text-only warning label, 40 to 49 percent of the subjects bid more for the cigarette packages with text-only warning label than they bid for packages with the graphic warning label. The subjects' revealed lower willingness to pay for packages with graphic warning labels is consistent with our DCE results that graphic warning labels will reduce cigarette consumption. However, the experimental auction data do not include direct measures of consumer cigarette choices or their choices to attempt to quit smoking. The results are uninformative about whether the mechanisms through which graphic warning labels affected consumer willingness to pay. Rousu et al. (2014) propose a model that assumes the graphic warning labels provide information but acknowledge the limitation that "our model of the value of information does not make a distinction between different types of information effects and whether they are based on analytical reasoning or emotions."

DCE studies of tobacco product choices are also published in inter-disciplinary public health journals, including journals focusing on tobacco such as *Tobacco Control* and the *Journal of Nicotine and Tobacco Research*.[1] These articles are not included in our Table C1 review because they have a different focus for a different audience. For example, Shang, Huang, Chaloupka, and Emery (2018) focus on the roles that flavors, e-cigarette device type, and e-cigarette warning labels play in youth stated preferences to try e-cigarettes. Subjects were presented with e-cigarette products with varying attributes but were not given the alternative of choosing combustible cigarettes.

Outside economics journals, the term "discrete choice experiment" is used in both a broad and narrow sense. In the broad sense, DCE has been used to describe various surveys that asks subjects to make choices with random assignment of the descriptions of the alternatives. In contrast, Louviere, Flynn, and Carson (2010) define DCEs as being necessarily grounded in random utility theory. Some of the studies published in public health journals do not fit this narrow definition of DCEs, even though the studies use the term to describe their research method. For example, Reynolds, Popova, Ashley et al. (2022) report a DCE about very low nicotine cigarettes (VLNCs) that asked respondents which message would most motivate them and least motivate them to quit smoking; the message attributes varied in terms of content about VLNCs and the source of the message. Subjects did not make choices between products. This is a study of consumer perceptions of message effectiveness which cannot be grounded in random utility theory. A related concern is that in some DCEs where subjects make choices between tobacco products, the product attributes are described in terms of consumer perceptions rather than observable characteristics of the products and/or the product marketplace. For example, Shang, Weaver, White, et al. (2020; not included in Table C1 because it is published in a public health journal) report a DCE about e-cigarettes that included "less harmful to health than cigarettes" and "effective for helping people quit" as product attributes. The Marti et al. (2019) study included in Table C1 also uses this approach. Although the results of these studies provide information about the relative importance

[1] Regmi, Kaphle, Timilsina, and Tuha (2018) report a systematic review of peer-reviewed studies published from 2000 – 2016 that used DCE methods in tobacco control. Of the 12 studies included in their review, 4 were published in health economics journals. Because these 4 studies focused on pharmaceutical smoking cessation products and did not include e-cigarettes, they are not included in Table C1.

of these perceptions on tobacco product choices, the results are less useful for policy analysis because perceptions are not directly policy manipulable. For this reason, in our DCE we describe the policy-manipulable attribute of warning labels.

Several studies published in public health journals use methods similar to DCEs to estimate the impact of different sizes of graphic warning labels (Kotnowski, Fong, Gallopel-Morvan et al. 2016; Barrientos-Guitierrez, Islam, Cho et al. 2021; Sillero-Rejon, Mahmoud, Tamayo et al. 2022). The studies were conducted in countries where graphic warning labels were already required. The studies did not show subjects packs with only text warning labels but instead used the currently required graphic warning label size as the control condition to be compared to packs with larger graphic warning labels. Each of these studies describes the method used as a DCE, but the methods are not grounded in random utility theory. Kotnowski et al. (2016) asked young smoking and non-smoking Canadian females about their interest in trying the different packs presented as choices, as well as which pack they thought would taste better and would be less harmful. Barrientos-Guitierrez et al. (2021) asked smoking and non-smoking early adolescents in Mexico to rank different packs in terms of attractiveness, likely to smoke, and healthiness. Sillero-Rejon et al. (2022) asked smoking and non-smoking adults aged 18 to 40 in Colombia about their preferences to try different packs and their perceptions of the taste and harm. The studies did not present subjects with the alternative of purchasing other products such as e-cigarettes. The studies also did not include an explicit option to not smoke or to quit smoking, although they were given the option to opt out and not choose any of the options presented. The studies provide evidence that larger graphic warning labels affect subjects' interest in cigarettes and their perceptions of taste and harm. However, they are not directly comparable to our study due to differences in study population, the lack of a control group shown text-only labels, and differences in DCE methods.

D. Empirical Evidence on the Internal and External Validity of SP Data Collected through the Cornell DCE

D.1. Validity Checks of the Quality of the Stated Preference Data

In sub-sections D.1 and D.2 we present empirical evidence on the internal and external validity of the SP data we collected through our DCE. Like other experimental research designs, the randomly assigned variation in product attributes in DCEs provides an internally valid research design to estimate the causal treatment effects of product attributes on subjects' stated preferences for tobacco products and quit attempts. However, because SP data are the subjects' responses about hypothetical choices, some subjects might not provide thoughtful and meaningful responses that provide useful information about the actual choices they would make in real-world markets. In this section D.1, we report the results of validity checks on the quality of our SP data and the implications for the empirical results reported in the text of the paper.

As an overview of the sensitivity of the empirical results to the validity checks, Table D1 reports the sensitivity of conditional logit models of tobacco product choices to alternative approaches to improve SP data validity. Column (0) reports a baseline conditional logit model estimated over the same sample used in estimation of the main text models. Columns (1) – (3) report conditional logit models estimated using sample restrictions to improve the quality of the SP data. Column (4) reports a conditional logit model estimated using a combination of SP and revealed preference

(RP) data. In columns (1) – (3) the point estimates of the alternative specific constants (ASCs) and the product attribute parameters tend to be very similar to the baseline model parameter estimates in column (0). The combined SP + RP model reported in column (4) yields estimated ASCs that are substantively different than the estimated ASCs in the baseline column (0) model. We will discuss the SP + RP data model results in more detail in the next sub-section D.2 of this Appendix.

In this sub-section we focus on sample restrictions that might improve the quality of our SP data. First, we examine data on the length of time subjects spent answering the DCE choice tasks, to identify possible "speedsters" who provided lower-quality responses. Figure D1 shows the distribution of time spent on the choice tasks. The median and mode times spent on the choice tasks are 4.3 and 3.4 minutes, respectively. Each subject completed 12 immediate choice tasks and 12 six-months-from-now choice tasks; the six-months-from-now choice tasks might be easier to complete quickly because they presented the subject with the same choices as in the preceding immediate choice task. For the model reported in column (1) of Table B1, we drop responses from 144 speedster subjects who completed the choice tasks in under 2 minutes.

Second, we examine data on the extent to which subjects paid attention to attribute variation across choice tasks. After subjects completed the choice tasks, we asked subjects which attributes varied across the tasks. Figure D2 shows the fraction of subjects who correctly indicated that the attribute in question varied. Although substantial fractions of the responses about attribute variation were incorrect, the results suggest that subjects paid the most attention to the price attribute of cigarettes, which was the most common tobacco product choice. The patterns of attentiveness across attributes and products are consistent with rational decisions to pay the most attention to the attributes and products that matter to their preferences. We also note that there is an ambiguity in our measure of attentiveness. As noted above, the six-months-from-now choice task was always identical to the preceding immediate choice task, i.e., in those pairs of tasks the attributes did not vary. For the model reported in column (2) of Table D1, we drop responses from 296 subjects who were inattentive to variation in the attributes of cigarettes.

Third, we examine data on the extent to which subjects might have made irrational choices across the choice tasks. Lancsar and Louivere (2006) discuss conventional approaches to rationality in consumer demand theory and the application to SP data collected through DCEs. Our first validity check examines whether a subject's choices across different choice tasks violates the weak axiom of revealed preference (WARP), with respect to differences in the price attributes. An example of a WARP violation with respect to the price attribute would be if a subject chooses e-cigarettes at a high price in one task, and in another task choose cigarettes even though the e-cigarette price is lower and the price of cigarettes is unchanged. Our second validity check examines WARP violations with respect to differences in the price attributes, the availability of flavors, and the availability of different levels of nicotine. An example of a WARP violation with respect to the price, flavor, and nicotine attributes would be if a subject chooses e-cigarettes at a high price in one task, and in another task chooses cigarettes even though the e-cigarette price is lower, and other e-cigarette attributes and the price of cigarettes are unchanged. Table D2 provides a summary of the criteria used to detect irrational choices and the fraction of irrational choices made by subjects in our graphic warning labels DCE. Considering the 12 immediate choice tasks, the average subject made 1.54 choices that violated WARP with respect to prices, and 0.86 choices that violated WARP with respect to the price, flavor, and nicotine attributes. The number of choices that violated WARP with respect to prices is higher because some of those choices might reflect the subject's willingness to make tradeoffs between price and the other attributes. For the model reported in column (3) of Table D1, we drop responses from the 309 subjects who made any choices that violated WARP with respect to the price, flavor, and nicotine attributes.

We note that our SP data include subjects who were non-traders and chose the same tobacco product, usually cigarettes, in all choice tasks. Lancsar and Louivere (2006) argue that although the responses from non-traders do not help identify marginal rates of substitution between attributes, dropping non-traders from the sample might eliminate subjects with strong preferences for a product or attribute.

D.2 Improving Data Quality by Combining SP and RP Data

In this sub-section we discuss the approach reported in column (4) of Table D1, where we estimate a conditional logit model of tobacco product choices using a combination of SP and RP data. In his monograph on econometric analysis of discrete choice data, Train (2002, pp. 174-175) discusses the advantages and disadvantages of SP and RP data:

> Revealed preference data have the advantage that they reflect actual choices…. However, RP data are limited to the choice situations and attributes of alternatives that currently exist or have existed historically. Often a researcher will want to examine people's responses in situations that do not currently exist, such as the demand for a new product. RP data are simply not available for these new situations.
>
> Stated-preference data complement revealed-preference data…. The limitations of SP data are obvious: what people say they will do is often not the same as what they actually do. People might not know what they would do if a hypothetical situation were real. Or they might not be willing to say what they would do.

Train suggests that by combining RP and SP data, "the advantages of each can be obtained while mitigating the limitations. The SP data provide the needed variation in attributes, while the RP data ground the predicted shares in reality." He outlines the approach we take, where we use our DCE's subjects' responses about their tobacco product use and quit attempts over the past year as SP data to calibrate our model.

To construct RP choices, we propose a probabilistic classification rule that uses information from the background survey of respondents' smoking behaviors. The process and classification rule are described in table D3. The construction takes two steps, first, we classify respondents as either choosing cigarettes or e-cigarettes. Second, we classify respondents as either choosing quit or not quit. To classify the choices of cigarettes and e-cigarettes, we use information of subjects' vaping history, vaping status, and smoking status. Among all the subjects, those who have never vaped are classified as choosing cigarettes, among those who have ever vaped, if they currently do not vape at all, they are classified as choosing cigarettes, for those who vape ever day and someday, we classify their choices according to their smoking status. Specifically, among those who vape ever day, if they also smoke every day, then their probabilities of choosing cigarettes versus e-cigarettes are 50% versus 50%, if they smoke someday (say m days out of 30 days), then their probabilities of choosing cigarettes versus e-cigarettes are $m/(30+m)$ versus $30/(30+m)$. For those who vape someday, the probability rules are similar, and the choice of quitting is constructed following a similar rule, details are reported in table D3.

After obtaining RP choices, we jointly estimate the model with SP and RP data. We assume that the random error term for the RP data follows an i.i.d. type I extreme value distribution with scale parameter normalized to 1, and the scale parameter for the SP data is given by λ. The choice probabilities for individual i chooses alternative j in the RP and SP data can then be written as:

$$P_{ij}^{RP} = \frac{e^{(\alpha_j + X_{ij}^{RP}\beta)}}{\sum_k e^{(\alpha_k + X_{ik}^{RP}\beta)}}$$

$$P_{ij}^{SP} = \frac{e^{[\lambda(\alpha_j + X_{ij}^{SP}\beta)]}}{\sum_k e^{[\lambda(\alpha_k + X_{ik}^{SP}\beta)]}}$$

In the joint estimation we maximize the joint likelihood function:

$$L(\alpha, \beta, \lambda) = \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij}^{RP} \log P_{ij}^{RP} + \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij}^{SP} \log P_{ij}^{SP}$$

The results are presented in column (4) of table D1. The estimated scale parameter λ is 0.6, suggesting that the scales in the RP and SP data are somewhat different. The estimated coefficients on the product attributes show similar patterns as in the column (0) model based on SP data only. We further use the model to predict the choice shares under a status quo scenario where cigarette prices equal to what the subjects paid in reality, e-cigarette prices are set to 8 dollars, e-cigarette available flavor includes tobacco, menthol, fruit/sweet/candy, e-cigarette available nicotine level is up to 50 mg, and e-cigarette warning message is the current FDA warning (contain nicotine, which is addictive). We conduct the prediction analysis with the without the presence of GWL separately, and by using the SP data only model and SP+RP data model separately. Results are presented in table D4, we see that the differences between the predicted choice shares with and without GWL roughly equal to the effects of GWL on choice probabilities we found in the paper, and the magnitudes predicted by SP data only model and SP+RP model are similar. Overall, we interpret the results from the calibrated model as supporting the usefulness of our SP data.

Table D1. Estimation Results from Conditional Logit Models

	Variables	Immediate Choice Today				
		(0)	(1)	(2)	(3)	(4)
Alternative specific constants	Cigarettes	1.933***	1.625***	2.450***	2.135***	3.285***
		(0.500)	(0.545)	(0.609)	(0.730)	(0.294)
	E-cigarettes	0.437	-0.041	0.879	-0.216	0.745***
		(0.544)	(0.602)	(0.679)	(0.821)	(0.234)
Cigarette warning	GWL amputation	-0.273***	-0.314***	-0.175	-0.365***	-0.494***
		(0.097)	(0.104)	(0.110)	(0.130)	(0.068)
Price	Price in USD	-0.101***	-0.104***	-0.128***	-0.121***	-0.172***
		(0.007)	(0.007)	(0.008)	(0.009)	(0.011)
		0.188***	0.209***	0.235***	0.319***	0.300***

E-cigarette available flavor	Tobacco, menthol, fruit/sweet/candy	(0.046)	(0.051)	(0.057)	(0.057)	(0.091)
	Tobacco and menthol	0.056	0.089*	0.123**	0.190***	0.077
		(0.043)	(0.048)	(0.054)	(0.056)	(0.099)
E-cigarette available nicotine level	Up to 20mg	0.017	0.000	-0.025	0.097**	0.007
		(0.040)	(0.044)	(0.048)	(0.046)	(0.073)
	Up to 50mg	-0.050	-0.053	-0.057	0.133***	-0.101
		(0.040)	(0.044)	(0.050)	(0.042)	(0.107)
E-cigarette warning	Are not completely risk free	-0.009	0.005	0.040	0.026	-0.033
		(0.046)	(0.050)	(0.055)	(0.057)	(0.107)
	Contain nicotine, which is addictive	-0.113**	-0.072	-0.136**	-0.025	-0.198**
		(0.056)	(0.060)	(0.067)	(0.069)	(0.099)
	May expose users to chemicals and toxins	-0.114**	-0.115**	-0.111*	-0.109*	-0.216**
		(0.048)	(0.054)	(0.059)	(0.057)	(0.131)
Scale parameter of SP data						0.600***
						(0.024)
Log-likelihood at convergence		-11428.869	-9838.842	-8150.374	-7545.087	-11904.921
Respondents		1171	1027	875	862	1171
Observations		42156	36972	31500	31032	42156

Notes: Additional control variables include smoking status (being every day smoker), age of smoking initiation, heavy smokers (heavy smoker index >= 3), vaping status (ever used e-cigarette), quitting effort (tried to quit during last 12 months), quit intention (intend to quit in the next 6 months), gender (being male), age, household size, education (reference category is grade school/some high school), household income above 50k, full-time employed, white. The reference category of e-cigarette available flavor “tobacco only”, the reference category of e-cigarette available nicotine level is “up to 5mg”, the reference category of e-cigarette warning is “no warning”. Standard errors are clustered at the individual level (except for column (4)). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

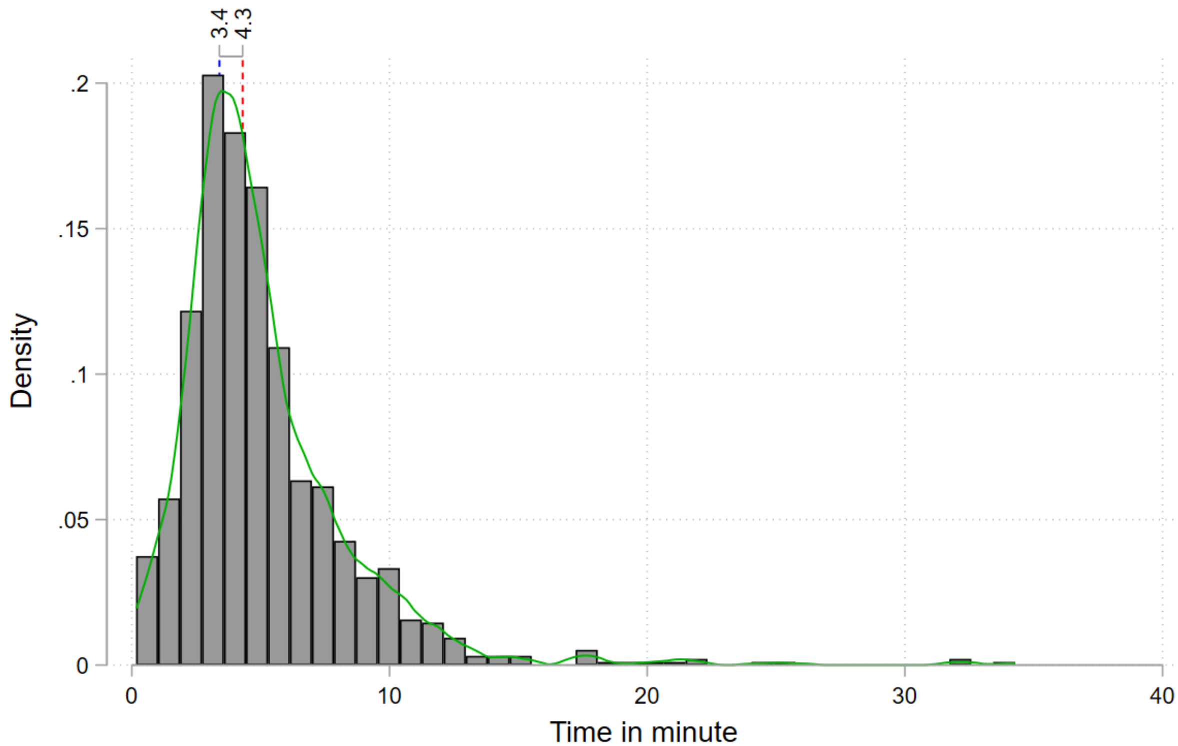

Figure D1. Distribution of Subject's Time Spent on the Choice Tasks

Notes: The blue and red dash line indicate the mode and median of time spent on the DCE section respectively. Out of 1171 subjects, 44 are dropped due to zero measured time spent on the choice tasks, the histogram represents the distribution of 1127 subject's time spent on the choice tasks.

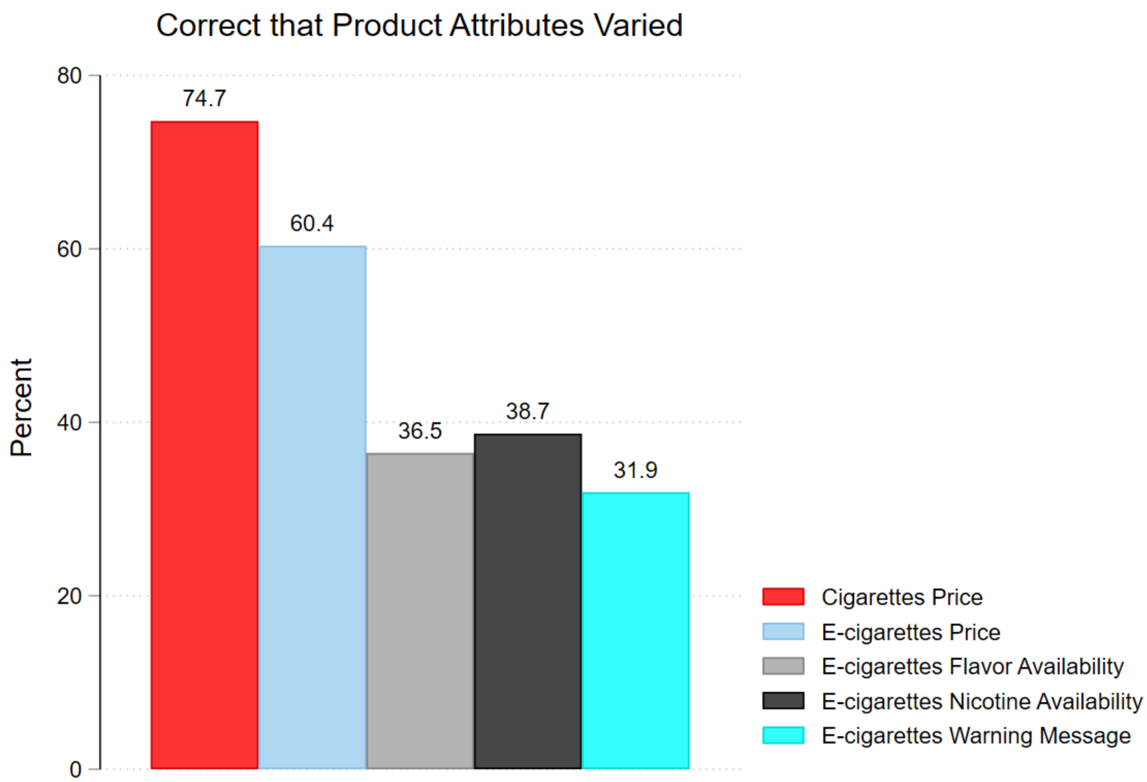

Figure D2. Fraction of Subjects that Were Correct that the Attribute in Question Varied.
Notes: Subjects are allowed to give correct but inconsistent answers, for example, they could select both price varied across scenarios and price was always the same. We use a strict criterion that the subject was correct that one attribute in question varied if she only selects the attribute varied and does not select the attribute was always the same.

Table D2. Criteria for Identifying Inconsistent Choices in the DCE

Choices at any two tasks		Current vs previous attributes[1]		Inconsistent?	% or # Inconsistent		
at previous task	at current task	Cigarettes[2]	E-cigarettes[3]			All attributes[3]	Prices only
					Task level: % (*N = 13,222)*		
Cigarettes	E-cigarettes	Same or better	Same or worse	Yes		2	4
Cigarettes	Quit	Better	(Any[4])	Yes		2	2
Cigarettes	Cigarettes	(Any)	(Any)	No			
E-cigarettes	Cigarettes	Same or worse	Same or better	Yes		2	4
E-cigarettes	Quit	(Any)	Better	Yes		0	1
E-cigarettes	E-cigarettes	(Any)	(Any)	No			
Quit	Cigarettes	Worse	(Any)	Yes		2	2
Quit	E-cigarettes	(Any)	Worse	Yes		0	0
Quit	Quit	(Any)	(Any)	No			
					Across all (6) rules	**8**	**14**
					Subject level: # per subject *(N = 1,202)*	**0.86** (range:0~10)	**1.54** (range:0~11)

Notes:

[1] Product attributes at the current task compared to the previous task.

[2] Cigarettes' attribute includes price only; Cigarettes' attribute becoming "better (worse)" means Cigarette price decreases (increases) at the current task compared to the previous task.

[3] E-cigarettes' attributes include price, flavor availability, and nicotine levels; E-cigarettes' attributes becoming "better (worse)" means E-cigarette price decreases (increases), flavor availability increases (decreases), and nicotine level increases (decreases) at the current task compared to the previous task. Warnings of E-cigarettes are not included in the criteria because subjects don't seem to have clear preference toward different warnings in the DCE.

[4] Product attributes at the current task can be the same, better, or worse compared to the previous task.

Table D3. A Probabilistic Classification Rule to Construct Revealed Preference Choices

Step 1: classification of choices between cigarettes and e-cigarettes					
Subject	Vaping history	Vaping status	Smoking status	Probability of choosing cigarettes	Probability of choosing e-cigarettes
N=1171	Vape ever (Yes) N=711	Vape every day (N=95)	Smoke every day (N=44)	50%	50%
			Smoke someday (m out of 30 days) (N=51)	$m/(30+m)$	$30/(30+m)$
		Vape someday (n out of 30 days) (N=255)	Smoke every day (N=147)	$30/(30+n)$	$n/(30+n)$
			Smoke someday (m out of 30 days) (N=108)	$m/(m+n)$	$n/(m+n)$
		Not at all (N=361)	Smoke every day (N=256)	100%	0
			Smoke someday (N=105)		
	Vape every (No) N=460		Smoke every day (283) (N=283)		
			Smoke someday (177) (N=177)		

Step 2: classification of choices of quitting			
Subject	Quitting intention	Quitting history	Probability of choosing quitting
N=1171	Plan to quit (Yes) N=770	Tried quitting (N=529) Didn't try to quit (N=241)	1/26
	Plan to quit (No) N=401	Tried quitting (N=60)	1/52
		Didn't try to quit (N=341)	0

Table D4. Predicted Choice Shares under Status Quo Scenario

Alternatives	Predicted choice shares under status quo					
	Model with SP data			Model with SP and RP data		
	With GWL	Without GWL	Difference	With GWL	Without GWL	Difference
Cigarettes	0.617	0.669	-0.052	0.7135	0.7809	-0.0674
E-cigarettes	0.183	0.158	0.025	0.1403	0.1078	0.0325
Quitting	0.200	0.172	0.028	0.1462	0.1113	0.0349

Notes: The status quo refers to a scenario where cigarette prices equal to what the subjects paid in reality, e-cigarette prices are set to 8 dollars, e-cigarette available flavor includes tobacco, menthol, fruit/sweet/candy, e-cigarette available nicotine level is up to 50 mg, and e-cigarette warning message is the current FDA warning (contain nicotine, which is addictive).

Appendix References

Barrientos-Gutierrez, Inti, Farahnaz Islam, Yoo Jin Cho , Ramzi George Salloum, Jordan Louviere, Edna Arillo-Santillán, Luz Myriam Reynales-Shigematsu, Joaquin Barnoya, Belen Saenz de Miera Juarez, James Hardin, and James F. Thrasher (2021). Assessing cigarette packaging and labelling policy effects on early adolescents: results from a discrete choice experiment. *Tobacco Control* 30: 505-514.

Blass, A. A., Lach, S., & Manski, C. F. (2010). Using elicited choice probabilities to estimate random utility models: Preferences for electricity reliability. *International Economic Review*, *51*(2), 421-440.

Buckell, J., & Hess, S. (2019). Stubbing out hypothetical bias: improving tobacco market predictions by combining stated and revealed preference data. *Journal of Health Economics*, *65*, 93-102.

Buckell, J., Hensher, D. A., & Hess, S. (2021). Kicking the habit is hard: A hybrid choice model investigation into the role of addiction in smoking behavior. *Health Economics*, *30*(1), 3-19.

de Bekker-Grob, E. W., Donkers, B., Bliemer, M. C., Veldwijk, J., & Swait, J. D. (2020). Can healthcare choice be predicted using stated preference data? *Social Science & Medicine*, *246*, 112736.

DeCicca, P., Kenkel, D., & Lovenheim, M. F. (2022). The economics of tobacco regulation: a comprehensive review. *Journal of Economic Literature*, *60*(3), 883-970.

Guindon, G. E., Mentzakis, E., & Buckley, N. J. (2024). "Cigarette packaging, warnings, prices, and contraband: A discrete choice experiment among smokers in Ontario, Canada." *Economics & Human Biology* 52: 101340.

Kenkel, D. S., Peng, S., Pesko, M. F., & Wang, H. (2020). Mostly harmless regulation? Electronic cigarettes, public policy, and consumer welfare. *Health Economics*, *29*(11), 1364-1377.

Kesternich, I., Heiss, F., McFadden, D., & Winter, J. (2013). Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D. *Journal of Health Economics*, *32*(6), 1313-1324.

Kotnowski, Kathy, Geoffrey T. Fong, Karine Gallopel-Morvan, Towhidul Islam, and David Hammond (2016). The Impact of Cigarette Packaging Design Among Young Females in Canada: Findings From a Discrete Choice Experiment. *Nicotine & Tobacco Research* 1348–1356.

Lancsar, E., & Louviere, J. (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Economics*, *15*(8), 797-811.

Linley, W. G., & Hughes, D. A. (2013). Decision-makers' preferences for approving new medicines in wales: a discrete-choice experiment with assessment of external validity. *Pharmacoeconomics*, *31*, 345-355.

Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling*, *3*(3), 57-72.

Maestas, N., Mullen, K. J., & Powell, D. (2023). The Effect of Population Aging on Economic Growth, the Labor Force, and Productivity. *American Economic Journal: Macroeconomics*, *15*(2), 306-332.

Marti, J., Buckell, J., Maclean, J. C., & Sindelar, J. (2019). To "vape" or smoke? Experimental evidence on adult smokers. *Economic Inquiry*, *57*(1), 705-725.

Mas, A., & Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, *107*(12), 3722-3759.

McFadden, Daniel (2017). "Stated Preference Methods and Their Applicability to Environmental Use and Non-use Valuations." *In Contingent Valuation of Environmental Goods: A Comprehensive Critique.* Daniel McFadden and Kenneth Train, editors. Edward Elgar Publishing. Elgaronline.com.

Mohammadi, T., Bansback, N., Marra, F., Khakban, A., Campbell, J.R., FitzGerald, J.M., Lynd, L.D., & Marra,C.A. (2017) Testing the External Validity of a Discrete Choice Experiment Method: An Application to Latent Tuberculosis Infection Treatment, Value in Health, 20 (7), 969-975.

Moshary, S., Shapiro, B., & Drango, S. (2023). *Preferences for firearms and their implications for regulation* (No. w30934). National Bureau of Economic Research.

Penn, J. M., & Hu, W. (2018). Understanding hypothetical bias: An enhanced meta-analysis. *American Journal of Agricultural Economics*, *100*(4), 1186-1206.

Quaife, M., Terris-Prestholt, F., Di Tanna, G. L., & Vickerman, P. (2018). How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. *The European Journal of Health Economics*, *19*, 1053-1066.

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, 5-17.

Reynolds, R. M., Popova, L., Ashley, D. L., Henderson, K. C., Ntansah, C. A., Yang, B., ... & Thrasher, J. (2022). Messaging about very low nicotine cigarettes (VLNCs) to influence policy attitudes, harm perceptions and smoking motivations: a discrete choice experiment. *Tobacco Control*.

Rousu, M.C., Marette, S., Thrasher, J. F., Lusk, J. L. (2014). The economic value to smokers of graphic warning labels on cigarettes: Evidence from combining market and experimental auction data. J*ournal of Economic Behavior and Organization,* 108, 123-134.

Ryan, M., Gerard, K., & Amaya-Amaya, M. (Eds.). (2007). *Using discrete choice experiments to value health and health care* (Vol. 11). Springer Science & Business Media.

Shang, C., Huang, J., Chaloupka, F. J., and Emery, S. L. (2018). The impact of flavor, device type, and warning messages on youth preferences for electronic nicotine delivery systems: evidence from an online discrete choice experiment. *Tobacco Control* 27: e152-e159.

Shang, C., Weaver, S. R., White, J. S., Huang, J., Nonnemaker, J., Cheng, K. W., & Chaloupka, F. J. (2020). E-Cigarette product preferences among adult smokers: a discrete choice experiment. *Tobacco Regulatory Science*, *6*(1), 66.

Sillero-Rejon, Carlos, Osama Mahmoud, Ricardo M. Tamayo, lvaro Arturo Clavijo-Alvarez, Sally Adams, and Olivia M. Maynard (2022). Standardised packs and larger health warnings: visual attention and perceptions among Colombian smokers and non-smokers. *Addiction* 17:1737–1747.

Telser, H., & Zweifel, P. (2007). Validity of discrete-choice experiments evidence for health risk reduction. *Applied Economics*, *39*(1), 69-78.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.