Examining the relationship between psychosocial and behavioral proxies for future consumption behavior: self-reported impact and bidding behavior in an experimental auction study on cigarette labeling

Matthew C. Rousu^{1*} and James F. Thrasher^{2,3}

¹Department of Economics, Susquehanna University, Selinsgrove, PA 17870, USA, ²Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, USA and

³Centro de Investigaciones en Salud Poblacional, Instituto Nacional de Salud Pública, Cuernavaca, Mexico

*Correspondence to: M. Rousu. E-mail: rousu@susqu.edu

Received on December 10, 2012; accepted on November 25, 2013

Abstract

Experimental and observational research often involves asking consumers to self-report the impact of some proposed option. Because self-reported responses involve no consequence to the respondent for falsely revealing how he or she feels about an issue, self-reports may be subject to social desirability and other influences that bias responses in important ways. In this article, we analyzed data from an experiment on the impact of cigarette packaging and pack warnings, comparing smokers' self-reported impact (four-item scale) and the bids they placed in experimental auctions to estimate differences in demand. The results were consistent across methods; however, the estimated effect size associated with different warning labels was two times greater for the four-item self-reported response scale when compared to the change in demand as indicated by auction bids. Our study provides evidence that self-reported psychosocial responses provide a valid proxy for behavioral change as reflected by experimental auction bidding behavior. More research is needed to better understand the advantages and disadvantages of behavioral economic methods and traditional self-report approaches to evaluating health behavior change interventions.

Introduction

Experimental and observational research on the impact of communications and marketing often involves asking consumers to report the impact and efficacy of various options. Such assessments are often theoretically based, focusing on psychosocial concepts that mediate the relationship between message exposure and downstream behavioral change, such as knowledge, attitudes and intentions [1, 2]. The use of this approach is common in public health research, with many studies using self-reports to evaluate the impact of food and cigarette labeling policies, novel tobacco products and online gambling problems [3–18]. Because these self-reported responses involve no consequence to the respondent for falsely revealing how he or she feels about an issue, they may be subject to social desirability and other demand characteristics that bias responses in important ways [19]. To better understand the implications of different methods for assessing product regulation impacts on consumer perceptions and behavior, this study uses data from experimental auctions and post-auction surveys to compare adult smokers' self-reported psychosocial and behavioral responses to different warning labels and brand elements on cigarette packaging.

Background

Conceptual models of how communications, including product packaging, influence behavior suggest that their effects on affective and cognitive variables are greater than on more distal behaviors [20–21]. In other words, most conceptual models posit that communication stimuli will have a weaker effect on behaviors than on the psychosocial variables that mediate the effect of the characteristic on the behavior. Furthermore, demand characteristics created by the data collection context may cause participants to over-report psychosocial responses. Hence, even if stimuli produce significant changes in self-reported perceptions or intentions, these changes do not necessarily translate into behavioral change. Some behavioral economics methods aim to address this issue by placing participants in situations with real financial consequences, thereby providing participants with enhanced incentives to provide truthful responses. For example, experimental auctions have been used for a wide array of topics, including the willingness-to-pay/willingness-to-accept disparity [22], the value of information about genetically modified foods [23] and how pictorial labels on cigarette packages affect demand [24, 25]

Increasingly, behavioral economics methods have been integrated into public health intervention development [26–28]. Research is needed to determine the consistency and implications of using these alternative approaches to estimating the behavioral impact of public health interventions. Behavioral economic research commonly assesses how demand for a product changes as a function of information transmitted through product packaging and promotional information. As such, the approach examines the relationship between product, promotions and price, which comprise three of the four ‘Ps’ used in social marketing approaches that have become increasingly popular with public health researchers and practitioners [29]. For communication interventions that involve product packaging, such as health-related labels, or provision of product information through other means, such as campaigns, behavioral economic methods provide a means for assessing their impact relative to price. For example,

this research approach is well-suited to inform Food and Drug Administration regulations over cigarette and food labeling and marketing [30], for which pre-market research to estimate effects has often relied on consumer self-reports ([31, 32] cites).

Economists who study non-market valuation (issues i.e., studies using non-market valuation attempt to place a value on items when a traditional market is not available) have written and published hundreds of studies using hypothetical ‘stated preference’ studies. These studies are used to value items where prices are not easily obtainable, and they cover a wide variety of topics, from the values of lakes and rivers [33] to the costs of the Exxon-Mobil oil spill [34]. Carson *et al.* [35, 36] list more than 1600 papers written using the contingent valuation method, which obtains hypothetical demand values from surveys by asking people about their ‘willingness to pay’ for a service or product. However, individuals generally inflate or overestimate their willingness to pay [37]. In other words, the self-reported values participants provide do not match behavior found in market settings.

The level of this bias has been the subject of numerous articles and debate. For example, a panel which contained two Nobel Prize winning economists recommended using a ‘divide by 2’ rule as a default rule [38, 39], indicating that when asked a willingness to pay question on a survey, respondents (on average) would claim they are willing to pay twice as much as they would actually be willing to pay. Fox *et al.* [37] found willingness to pay responses were ~50% higher than auction market bids for sandwiches but also determined that the type of product mattered. The level of bias was larger for a more controversial product (irradiated sandwich) than a less controversial product (regular sandwich). Other studies have also found survey responses did not match behavioral changes in the market. For example, List and Shogren [40] examined the gap between a person’s ‘willingness to accept’ compensation to be harmed and found that what participants claimed they would need (hypothetically) was different than what they actually were willing to take. Other studies, while all acknowledging that substantial differences between

self-reported responses and behavior have attempted to quantify the differences [41–44].

Experimental auction methods aim to overcome some of the limitations of ‘willingness to pay’ methods. In 2002, Vernon Smith won the Nobel Prize in Economics for refining experimental economics methods [45, 46]. A key benefit of experimental auctions comes from placing consumers in a real auction, with winners and losers, and where the winners ‘pay’ for products. This prevents hypothetical bias of ‘willingness to pay’ studies by providing financial consequences for self-reported valuation of a product. In experimental auctions, consumers assign monetary value to a product as an estimate of product demand. Since their bids are on a monetary scale, the variability of options potentially allows for greater sensitivity than standard self-reports of intention to buy. Hence, bids for the same product under different conditions can be used to estimate differences in product demand between conditions. This approach is ideal for estimating the changes in demand that could stem from implementing packaging and labeling policy alternatives. Recently, experimental auctions have been used to assess smokers’ demand for low and no-nicotine products [47], for potentially reduced exposure products [48], foods sold with different kinds of labels [23, 49].

This study aimed to determine the consistency of self-reported psychosocial effects and auction bids estimating demand changes when adult smokers evaluated cigarette packs with different size and formats of health warning labels (HWLs), as well as when the cigarette packs contained no brand imagery or colors. This study is a secondary analysis of data from Thrasher *et al.* [25] and Rousu and Thrasher [50]. We expect labeling and packaging effects to be stronger for self-reported impact than for auction bids.

Methods

Experimental design

The data for this study are from Thrasher *et al.* [25], where a full description of the experimental design

is presented. The study protocol was approved by the IRB at the University of South Carolina. To recruit participants, tables were set up at grocery stores in four cities: Selinsgrove, PA; Columbia, SC; Tampa, FL; and San Diego, CA between May and September 2009. Eligible study participants were 18 and older, had smoked more than 100 cigarettes in their lifetimes, had smoked at least one cigarette in the last month and were not pregnant. Posted signs indicated that adult smokers could earn \$15 for 15 minutes of their time. This money not only was enough to compensate them for their time but also ensured participants had money to place bids (and potentially purchase) products in the auction. Auctions were conducted with one to eight participants at a time, and a total of 402 participated, with each participant bidding on two different packs of cigarettes. Table I presents the characteristics of our sample.

The study involved assessing four HWL conditions (see Fig. 1), all with the same novel message (i.e. ‘Smoking causes mouth cancer’), which is not currently on US HWLs: 1. text-only message that covered 50% of one side of the package (current US policy); 2. text-only message that covered 50% of the lower half of the front, back and one side of the package; 3. text message with pictorial image of mouth cancer, covering 50% of the lower half of the front, back and one side of the package; 4. the same text and pictorial image as in condition 3, but with color and symbolic brand elements removed, while maintaining the brand font, size and descriptors. Participants evaluated branded packs that are most popular within the three main tobacco product classes (i.e. Marlboro Red = full flavor smokers; Marlboro Light = light smokers; Newport = menthol smokers) and were randomly assigned to different treatments that involved evaluating two of the four different packages, with type and order counter-balanced. For this study, we only report on the first package participants viewed and bid upon to avoid issues with ordering effects and anchoring that might occur both with bids and with the survey responses [51, 52].

Participants first filled out a brief questionnaire on smoking behavior, after which study personnel

Table I. Characteristics of study sample and associated experimental conditions

	Control text (<i>N</i> = 98)	Larger text (<i>N</i> = 140)	Picture and text (<i>N</i> = 104)	Picture, text and plain package (<i>N</i> = 60)	Entire sample (<i>N</i> = 402)
Age (mean)	38.2	37.4	39.0	37.3	38.0
Female	49%	45%	44%	42%	44%
Race/ethnicity					
White	65%	57%	61%	53%	59%
Black	31%	35%	37%	43%	36%
Other	4%	7%	2%	4%	5%
Annual household income					
Less than \$15 000	42%	47%	45%	45%	43%
\$15 000–\$35 000	28%	38%	33%	35%	35%
More than \$35 000	30% ^a	15%	22%	21%	18%
Educational attainment					
Less than HS	18%	14%	15%	16%	16%
HS degree but no college	35%	46%	44%	42%	41%
At least some college	47%	44%	41%	42%	43%
Number of cigarettes smoked per day	16.0	17.2	17.0	14.9	16.5
% wishing to quit within 6 months	47%	60%	58%	57%	56%

^aParticipants in the control group had significantly higher income than other groups, at $P < 0.05$.

explained the auction (see below), including why it was in their best interest to bid actual value that they ascribed to the products, no more and no less. Any questions they had were answered and a practice auction round was conducted with a candy bar before proceeding with the auction.

Measurement

Demand. Participant bids for cigarette packages with different HWLs were collected using the demand revealing Becker–DeGroot–Marschak (BDM) auction mechanism [53]. Consistent with this approach, each participant was presented with and asked to bid on a cigarette pack. Once this was done, participants were presented with the second pack, and they bid on it. Next, the binding auction round (i.e. which pack would be auctioned) was randomly determined. Only one round was chosen as binding to avoid decreasing bids from demand curve effects [54]. The selected price was then randomly chosen from a uniform distribution, which ranged from \$0.10 to \$10.00 in increments of \$0.10. If the participant bid more than this randomly selected price, she paid the selected price and received the package.

According to microeconomic theory, a bid obtained from the BDM mechanism represents a point on the consumer's demand curve [55]. The BDM mechanism is demand revealing, as the bid accurately represents a consumer's demand for a product [55]. Participants have an incentive to bid their actual value for the product, as bids that are higher than this value may result in paying too much for the product, whereas bids that are lower could result in their not obtaining the product at a favorable price (e.g. see Corrigan and Rousu [56] or Thrasher *et al.* [25]). The benefit of using this auction mechanism is that bidding behavior (i.e. the bid price) captures consumer behavior, as long as it is reasonable to assume that the person who places the bid is the person who consumes the product, which in our case is very likely given that smokers are bidding on cigarettes.

This type of experimental mechanism has been shown to have validity in predicting consumer choices in the marketplace (i.e. have external validity). Chang *et al.* [57] tested both hypothetical and non-hypothetical mechanisms and found that a non-hypothetical experiment similar to the approach used here outperformed hypothetical mechanisms

Fig. 1. Labels used in the experiment.

and did a good job of predicting retail sales. Ding *et al.* [58] showed that bids from experimental auctions predicted non-hypothetical choices in external environments.

Self-reported characteristics and perceptions of warnings. After participants viewed the cigarette package with the HWL and placed a bid on it, we asked participants four questions used in other studies of the perceived impact of the HWLs:

- ‘Please tell me how much this health warning makes you think about the dangers of smoking’,
- ‘How much does this health warning make you think about quitting smoking’,
- ‘How much do you think this health warning would make other smokers think about quitting smoking’ and
- ‘How much do you think this health warning would keep children from starting to smoke?’

Response options included ‘Not at all’, ‘A little’, ‘Somewhat’ or ‘A lot’, and Cronbach’s alpha for these four items was 0.75. We intended to remove any item that decreased alpha (none did). Hence, we summed the scores to derive a scale for perceived HWL impact (the Table AI presents the mean estimates for each of the four questions).

Analysis

Data were analyzed using SAS statistical software and the analytic sample for this study involved observations from bids on and reported responses to the first pack of cigarettes to which participants were exposed. First, we assessed differences in demographic characteristics across experimental groups using chi-square tests and *t*-tests. Next, we estimated the mean bids and mean perceived impact for each condition and used a series of independent *t*-tests to assess the difference between means. The regression models allow us to estimate if any independent variables impacted bids or the perceived impact of warnings and if the same independent variables were statistically significant across models. For both these models, the treatment condition was entered as a dummy variable, with the current

warning label condition as the reference group. These analyses were also run after combining just the first two indicators (how much does this warning make you think about the dangers of smoking; how much does the warning make you think about quitting smoking), as they reflected smokers’ assessments of personal impacts. The results were similar and are available from the authors upon request.

Results

Table I shows the characteristics of the overall sample, as well as of the samples within each cigarette package bid condition. The mean age of participants was 38.0 years old, and 44% of the sample was female. Fifty-five percent of the sample was White, 41% Black and 5% identified as a different ethnic or racial background. Almost half (45%) of the sample had household incomes below \$15 000 and 60% had a high school degree or less. Study participants smoked an average of 16.5 cigarettes a day, and 56% indicated they were either currently trying to quit or planning to quit smoking within the next 6 months. No statistically significant differences were found in these characteristics across the four treatment conditions, except for the income variable, where participants in the control group had higher income than participants in the other groups.

Table II presents unconditional means for bids and perceived efficacy for each of the four HWLs. Both the experimental auction bids and efficacy scores show the same pattern: participants placed lower bids and had higher ratings of perceived impact for packs with pictorial HWLs. This is also true for each of the four individual questions on the perceived impact of HWLs (see Table AI), although the magnitudes of the effect appear to be different for each question, with the smallest effects apparent for the question on personal motivation to quit. The differences in mean bids were only statistically significant at the 10% level. The direction and magnitude are similar to our results in Thrasher *et al.* (2011), but aren’t as statistically significant, potentially due to lack of power with using half

Table II. *Bids and perceived efficacy across the experimental conditions*

	Control text (<i>N</i> = 98)	Larger text (<i>N</i> = 140)	Picture and text (<i>N</i> = 104)	Picture, text and plain package (<i>N</i> = 60)	Entire sample (<i>N</i> = 402)
Mean bid (SD) ^a	\$3.52 (1.59)	\$3.53 (1.46)	\$3.18 (1.56)	\$3.14 (1.55)	\$3.38 (1.54)
Mean total scale score ^b (SD)	9.49 (3.79)	10.4 (3.24)	11.57 (3.26)	11.6 (3.46)	10.66 (3.51)
% change in bid relative to control text ^a	0.0%	0.0%	−9.7%	−10.7%	
% change in total scale relative to control text ^b	0.0%	10.4%	21.9%	22.2% ^c	

^aDifferences between the bids across treatments are not statistically significant at the 5% level using a two-sided *t*-test.

^bThe differences in Likert-type responses across the four treatments are all statistically significant from each other at the 5% (or 1%) level using a two-sided *t*-test except for the difference between the plain packaging treatment and the picture and text treatment. Note: non-parametric Wilcoxon Rank-Sum test yielded similar results.

^cThe difference between the percentage increase in the total scale score and the percentage increase in bids is statistically significant at the 1% level using a two-sided *t*-test.

Table III. *Ordinary least squares regression models, predictors and correlates of bids for cigarettes with alternative warning labels and perceived efficacy (i.e. correlates of perceived effectiveness) of cigarette warning labels (*N* = 402)*

Characteristics	Dependent variable = bid	Dependent variable = scale score
	Multivariate model	Multivariate model
Experimental condition		
Intercept (i.e. control text only)	4.19 ^a (0.32)	6.21 ^a (0.59)
Large text versus control	0.03 (0.20)	0.54 (0.45)
Large text and picture versus control	−0.35 (0.22)	2.02 ^a (0.48)
Large text, picture and plain packaging versus control	−0.41 (0.25)	1.75 ^a (0.56)
Age	−0.02 ^a (0.01)	0.00 (0.01)
Female	−0.21 (0.16)	0.18 (0.34)
Race/ethnicity		
Black versus White	−0.05 (0.17)	0.78 ^b (0.38)
Other versus White	−0.73 ^b (0.33)	−0.18 (0.74)
Annual household income		
\$15 000–\$35 000 versus <\$15 000	−0.06 (0.18)	−0.60 (0.39)
More than \$35 000 versus <\$15 000	0.04 (0.22)	−0.76 (0.50)
Educational attainment		
Less than HS versus College +	0.21 (0.24)	−0.62 (0.54)
HS degree but no college versus College +	0.08 (0.18)	0.18 (0.39)
Number of cigarettes per day	0.02 ^a (0.01)	−0.02 (0.01)
Intend to quit within the next 6 months	−0.06 (0.16)	1.43 ^b (0.34)

^a*P* < 0.01; ^b*P* < 0.05.

observations as Thrasher *et al.* (2011). The differences in the self-reported response scale were statistically significant at the 5% or 1% level. Participants bid 10.7% less for cigarettes with the pictorial HWL and a plain label, while there was a 22.2% increase in the perceived impact scale. This represents a 107% greater change in the self-reported

psychosocial responses than with the bid prices. The difference between the 22.2% change in the perceived impact scale and 10.7% change in bids is statistically significant at the 1% level using a two-sided *t*-test.

Table III shows results from the regression models for bids and perceived impact. With regard

to differences in the evaluation of the warnings, the results are consistent with what was found above. Income and education were unassociated with either bids or perceived impact. However, statistically significant predictors differed across these two outcomes. Participants who were relatively older placed lower bids, while those who identified as other versus White ethnicity, and who smoked relatively more cigarettes per day were more likely to place higher bids. Participants were more likely to perceive warnings as effective if they identified as Black compared to White and if they intended to quit. For the regressions, we also ran models that included interactions between label condition and education, smoking intensity and intention to quit. None of the interaction terms were statistically significant at the 5% level. Those results are not reported here, but are available from the authors upon request.

Discussion

Assessing the impact of product packaging information on purchasing behavior is important to public health researchers and regulators who aim to promote consumption of healthy products. Also important to public health researchers is assessing how different ways of presenting health information influence psychosocial characteristics, including attitudes and intentions. Indeed, if the intent of regulatory policy is to inform consumers rather than change their behavior, then psychosocial outcomes are critical. However, if the policy intent is to change behavior then understanding how psychosocial assessments relate to demand changes will be useful to researchers.

Our study collected bids from experimental auctions, which estimate demand and therefore current consumption behavior. We also obtained survey responses on several Likert-type questions to examine perceived effectiveness of labels. The results from our study suggest that bids in an experimental auction and self-reported perceived impact produced relatively consistent responses across alternative warning labels for cigarettes—labels

with pictorial elements were associated with both reduced demand for cigarettes and higher perceived impact of the label.

As hypothesized, differences in self-reported, perceived impact of HWLs across conditions were greater than for auction bids (range of up to 22% versus up to 11%). This finding is consistent with economics research that has found hypothetical responses differ from non-hypothetical responses in willingness to pay studies [38, 41]. Self-reported psychosocial responses may inflate the magnitude of impact if the outcome of interest is behavior change. Nevertheless, most psychosocial theories place these psychosocial impacts along the pathway to behavior, while recognizing that they are not perfectly correlated with behavior (1, 21). Hence, some of the discrepancies around self-reported psychosocial impact and demand may be diminished by the time it translates to behavior. In that way, our results are consistent with the principle that psychosocial changes may be greater than behavioral changes. In this case, the behavior change is the change in demand for cigarettes. Furthermore, participants were asked to bid on the whole product, but the focal stimuli for the psychosocial questions were focused on the warning.

The consistency of results produced by differing methods provides evidence for using either approach when assessing impacts of labeling policy alternatives. More research is needed to verify further this finding across other communication stimuli and situations, including real world impacts of warning label policy implementation; nevertheless, this study suggests that self-reported psychosocial responses and bids produced relatively consistent responses. In addition, we did not find any interactions between treatments and key socio-demographic and smoking-related variables, indicating that while some groups may be willing to bid higher or provide higher self-reported scores, these differences did not appear to vary across treatment conditions.

The consistency of our results across methods is limited, however. Our results suggest researchers should use caution when interpreting magnitude of the changes either based only on bids or only on

self-reported measures. We found that effects on demand were less than changes in psycho-social responses. We find no evidence that statistically significant demographic characteristics for psychosocial responses would translate into bid changes. Given that independent variables which were statistically significant in one model were not in the other, researchers should be cautious how they interpret results found based on psychosocial responses, particularly if they suggest that these responses will translate into behavioral change. Change in the upstream factors such as perceptions of harm and quitting thoughts do not necessarily lead to behaviour change because there are many factors beyond the scope of labelling and packaging, such as addiction level and readiness to quit, that determine whether these changes lead to behavior change. More powerful cross-validation research is needed to directly compare self-reported behavioral intentions and bids with each other, as well as an assessment of their predictive validity under naturalistic conditions of policy or program implementation.

Our results indicated that Black smokers indicated greater effectiveness of labels compared to White smokers, which is consistent with the results from Thrasher *et al.* [15]. We have no explanation for this effect, but it is worth noting that Black smokers did not bid less for cigarettes in the auction. We find those in the 'other' race category, i.e. non-White, non-Black smokers, bid less for cigarettes but did not have statistically significant differences in their perceived efficacy. Age was negatively correlated with bid prices, but had no statistically significant effect on perceived efficacy. This is consistent with the idea that those who are older may be more price sensitive regarding cigarette purchases [59]. The number of cigarettes per day had positive effect on bids but not perceived efficacy, while the intention to quit in the next 6-month had a positive effect on perceived efficacy, which has been found in other research [15, 16], but not on bids. These two findings deserve elaboration. The number of cigarettes smoked per day is an indicator of addiction, and the most consistent predictor of successful smoking cessation [60]. Intention to quit predicts quit attempts and is a more consistent

predictor of self-reported efficacy of warnings than level of addiction; however, the relationship between intention to quit and quit success is much less consistent than has been found for level of addiction [60]. The differences in our findings may partly reflect this distinction. Purchase tasks may perform better than psychosocial assessments if the goal is to produce longer-term behavioral change. In addition, bidding was done with the warning label presented on the pack, which provides a cue to smokers that is independent of the warning. The warning label assessments, on the other hand, asked participants to focus on the warnings, even though they were presented on the pack. This may also account for the differential findings and their relationship to addiction level and intentions to quit.

There are several limitations of our study. Our behavioral assessment of demand and the 'perceived effectiveness' construct could have been more conceptually isomorphic, as divergence in the conceptual domain that we measured may also help explain divergent results. Our measure of perceived effectiveness included two items that assessed projected effects of the warning on future behavior (quit intention), which shares significant conceptual overlap with demand. The other two questions included in our scale assessed the influence of the warning on risk perceptions, which are proximal causes of behavioral intention according to standard theories used in communication research [1]. We combined these indicators because of their conceptual proximity, their high level of correlation and because of our desire to enhance variation in the psychosocial measure. Future research in this area should include measurement of psychosocial constructs that have a greater conceptual overlap with demand, including those that would assess changes in consumption that are more subtle than quitting altogether.

A second potential shortcoming of our study is that the response options for the ratings were limited to four options and we assigned points to responses that may not adequately reflect the strength of responses. Future research that examines bids versus self-reported responses should consider response options that represent a greater range or options (e.g. a 7–10 point scale) than we

used here. Nevertheless, a four-point scale is often used, so this shortcoming may actually be an additional caution for researchers when interpreting Likert-type responses with limited options. Third, the choice of exactly what psychosocial variables to include in a scale scores to include could lead to different results. For example, one of the four responses, on likelihood to quit, had a smaller percentage change than bid change. Furthermore, only two of the four variables specifically addressed behavior change. Further research should examine the consistency of psychosocial reports using more behavioral response questions. A fourth limitation is the order of eliciting bids first and self-reported responses afterwards may have biased results, as participants may have aligned their self-reported psychosocial impact assessments with their previous bids. This may have resulted in a greater correspondence between bids and self-reports than if the order of eliciting responses had been randomized. Future experiments could assess the ordering impact on the relative increase across methods.

Funding

Funding for data collection and analysis was received from the Robert Wood Johnson foundation grant titled “Estimating the impact of alternative warning labels on cigarette demand in the United States: Evidence from experimental auctions”, grant number RWJF 22-00-25528. Dr. Thrasher’s contribution to manuscript writing was partly supported by a grant from the National Cancer Institute (R01 CA167067; *Building evidence for effective and sustainable cigarette warning label policy*).

Conflict of interest statement

None declared.

References

1. Institute of Medicine. *Speaking of Health: Assessing Health Communication Strategies for Diverse Populations*.

- Washington, DC: Institute of Medicine, National Academies of Medicine, 2002.
2. National Cancer Institute. *Making Health Communications Work: A Planner’s Guide*. Washington, DC: US Department of Health and Human Services, National Institutes of Health, National Cancer Institute, Office of Communications, 2004.
3. Berg C, Thrasher J, Westmaas J *et al*. College student reactions to health warning labels: sociodemographic and psychosocial factors related to perceived effectiveness of different approaches. *Prev Med* 2011; **53**: 427–40.
4. Lund KE, McNeill A, Scheffels J. The use of snus for quitting smoking compared with medicinal products. *Nicotine Tob Res* 2010; **12**: 817–22.
5. Shadel WG, Martino SC, Haviland A *et al*. Smoking motives in movies are important for understanding adolescent smoking: a preliminary investigation. *Nicotine Tob Res* 2010; **12**: 850–4.
6. Terry-McElrath Y, Wakefield M, Ruel E *et al*. The effect of antismoking advertisement executional characteristics on youth comprehension, appraisal, recall, and engagement. *J Health Commun* 2005; **10**: 127–43.
7. Bansal-Travers M, Hammond D, Smith P *et al*. The impact of cigarette pack design, descriptors, and warning labels on risk perception in the U.S. *Am J Prev Med* 2011; **40**: 674–82.
8. Wakefield M, Durrant R, Terry-McElrath Y *et al*. Appraisal of anti-smoking advertisement by youth at risk for regular smoking: a comparative study in the United States, Australia, and Britain. *Tob Control* 2003; **12**(Suppl. II): ii82–6.
9. White V, Webster B, Wakefield M. Do graphic health warning labels have an impact on adolescents’ smoking-related beliefs and behaviors? *Addiction* 2008; **103**: 1562–71.
10. Hammond D, Fong GT, Borland R *et al*. Text and graphic warnings on cigarette packages: findings from the international tobacco control four country study. *Am J Prev Med*, 2007; **32**: 202–9.
11. Hammond D, Thrasher JF, Reid JL *et al*. Perceived effectiveness of pictorial health warnings among Mexican youth and adults: a population-level intervention with potential to reduce tobacco-related inequities. *Cancer Causes Control* 2012; **23**: 57–67.
12. Borland R, Yong H-H, Wilson N *et al*. How reactions to cigarette packet health warnings influence quitting: findings from the ITC four-country survey. *Addiction* 2009; **104**: 669–75.
13. Smith M, Rousu M, Dion P. Internet poker: examining motivations, behaviors, outcomes, and player traits using structural equations analysis. *J Gambl Issues* 2012; **27**: 1–23.
14. Thrasher JF, Arillo-Santillán E, Villalobos V *et al*. Can pictorial warning labels on cigarette packages address smoking-related health disparities? Field experiments in Mexico to assess pictorial warning label content. *Cancer Causes Control* 2012a; **23**: 69–80.
15. Thrasher JF, Carpenter M, Andrews JO *et al*. Cigarette warning label policy alternatives and smoking-related health disparities. *Am J Prev Med* 2012b; **43**: 590–600.
16. Thrasher JF, Pérez-Hernández R, Arillo-Santillán E *et al*. [Towards informed tobacco consumption in Mexico: effects of pictorial warning labels among smokers]. *Rev Salud Pública México* 2012c; **54**: 242–53.

17. Feunekes GIJ, Gortemaker IA, Willems AA *et al.* Front-of-pack nutrition labeling: testing effectiveness of different nutrition labeling formats front-of-pack in four European countries. *Appetite* 2008; **50**: 57–70.
18. Feunekes GIJ, de Graaf C, Meyboom S *et al.* Food choice and fat intake of adolescents and adults: associations of intakes within social networks. *Prev Med* 1998; **27**: 645–56.
19. Stone AA, Bachrach CA, Jobe JB *et al.* *The Science of Self-report: Implications for Research and Practice*. Mahwah, NJ: Lawrence Erlbaum, 2000.
20. Jackson C. Behavioral science theory and principles for practice in health education. *Health Educ Res* 1997; **12**: 143–50.
21. Klimchuk MR, Krasovec SA. *Packaging Design: Successful Product Branding from Concept to Shelf*. NJ: John Wiley & Sons, 2006.
22. Shogren J, Shin S, Hayes D *et al.* Resolving differences in willingness to pay and willingness to accept. *Am Econ Rev* 1994; **84**: 255–70.
23. Rousu M, Huffman WE, Shogren JF *et al.* Effects and value of verifiable information in a controversial market: evidence from lab auctions of genetically modified food. *Econ Inquiry* 2007; **45**: 409–32.
24. Thrasher JF, Rousu M, Ocampo-Anaya R *et al.* Estimating the impact of graphic warning labels on cigarette packs: the auction method. *Addict Behav* 2007; **32**: 2916–25.
25. Thrasher JF, Rousu M, Hammond D *et al.* Estimating the impact of Food and Drug Administration regulation of cigarette package warning labels and the added impact of plain packaging: evidence from experimental auctions among adult smokers. *Health Policy* 2011; **102**: 41–8.
26. Baicker K, Congdon WJ, Mullainathan S. Health insurance coverage and take-up: lessons from behavioral economics. *Milbank Q* 2012; **90**: 107–34.
27. Zimmerman 2009.
28. Cherukupalli 2010.
29. Grier S, Bryant CA. Social marketing in public health. *Annu Rev Public Health* 2005; **26**: 319–39.
30. Waxman H. Family Smoking Prevention and Tobacco Control Act. In: Congress tU (ed.). *H.R.1256*, 2009.
31. Nonnemaker J, Farrelly MC, Kamyab K *et al.* *Experimental Study of Graphic Cigarette Warning Labels*. Research Triangle Park, NC: RTI International, 2010.
32. Institute of Medicine. *Front-of-Package Nutrition Rating Systems and Symbols Promoting Healthier Choices: Promoting Healthier Choices*. Institute of Medicine. Washington DC: National Academies Press, 2012.
33. Corrigan JR, Egan KJ, Downing JA. Aesthetic values of lakes and rivers. In: Likens G (ed). *Encyclopedia of Inland Waters*.
34. Carson RT, Mitchell RC, Hanemann M, Kopp RJ, Presser S, Ruud PA. *A Contingent Valuation Study of Lost Passive Use Values Resulting From the Exxon Valdez Oil Spill*, 1992. Report to the Attorney General of the State of Alaska.
35. Carson RT, Wilks L, Imber D. Valuing the preservation of Australia's Kakadu Conservation Zone. *Oxford Econ Pap* 1994; **46**: 727–49.
36. Carson RT, Wright J, Carson N, Alberini A, Flores N. *A Bibliography of Contingent Valuation Studies and Papers*. La Jolla, CA: Natural Resource Damage Assessment, Inc, 1995.
37. Fox JA, Shogren JF, Hayes DJ *et al.* CVM-X: calibrating contingent values with experimental auction markets. *Am J Agric Econ* 1998; **80**: 455–65.
38. National Oceanic and Atmospheric Administration. Natural Resource Damage Assessment: Proposed Rules. *Federal Register* 4 May 1994, 23098–111.
39. National Oceanic and Atmospheric Administration. Natural Resource Damage Assessment: Final Rules. *Federal Register* 5 May 1996, 439.
40. List JA, Shogren JF. Calibration of willingness-to-accept. *J Environ Econ Manage* 2002; **43**: 219–33.
41. Murphy JJ, Stevens TH, Yadav L. A comparison of induced value and home-grown value experiments to test for hypothetical bias in contingent valuation. *Environ Resource Econ* 2010; **47**: 111–23.
42. Alfnes F, Yue C, Jensen HH. Cognitive dissonance as a means of reducing hypothetical bias. *Eur Rev Agric Econ* 2010; **37**: 147–63.
43. List JA. Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. *Am Econ Rev* 2001; **91**: 1498–507.
44. Cummings RG, Taylor LO. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *Am Econ Rev* 1999; **89**: 649–65.
45. Smith VL. An experimental study of competitive market behavior. *J Polit Econ* 1962; **70**: 11–137.
46. Smith VL. Experimental auction markets and the Walrasian hypothesis. *J Polit Econ* 1965; **73**: 387–93.
47. Monchuk D, Rousu M, Shogren J *et al.* Decomposing the value of cigarettes using experimental auctions. *Nicotine Tob Res* 2007; **9**: 93–9.
48. Rousu M, Nonnemaker J, Farrelly M. The value of counter-marketing information to smokers: evidence from field auctions. *J Agric Appl Econ* 2011; **43**: 607–20.
49. Lusk JL, Fox JA, Schroeder TC *et al.* In-store valuation of steak tenderness. *Am J Agric Econ* 2001; **83**: 539–50.
50. Rousu, Thrasher 2012.
51. Nunes, Boatwright 2004.
52. Corrigan *et al.* 2011.
53. Becker, DeGroot, Marschak 1964.
54. Rousu, Beach, Corrigan 2008.
55. Lusk, Shogren 2007.
56. Corrigan J, Rousu M. Are experimental auctions demand revealing when values are affiliated? *Am J Agric Econ* 2011; **93**: 514–20.
57. Chang JB, Lusk JL, Bailey Norwood F. How closely do hypothetical surveys and laboratory experiments predict field behavior? *Am J Agr Econ* 2009; **91**: 518–34.
58. Ding M, Grewal R, Liechty L. Incentive-aligned conjoint analysis. *J Marketing Res* 2005; **42**: 67–83.
59. Chaloupka FJ, Pacula RL. The impact of price on youth tobacco use. *Smoking and Tobacco Control Monographs: Monograph 14: Changing Adolescent Smoking Prevalence*. Bethesda, MD: National Cancer Institute, National Institutes of Health, USDHHS, 2001, 193–9.
60. Vangeli E, Stapleton J, Smit ES *et al.* Predictors of attempts to stop smoking and their success in adult general population samples: a systematic review. *Addiction* 2011; **106**: 2110–21.

Appendix

Table AI. *Scores for each individual question across the experimental conditions*

	Control text (<i>N</i> = 98)	Larger text (<i>N</i> = 140)	Picture and text (<i>N</i> = 104)	Picture, text and plain package (<i>N</i> = 60)	Entire sample (<i>N</i> = 402)
Label makes smoker think of the dangers of smoking (mean)	2.32	2.64 (13.8%)	2.88 (24.1%)	2.98 (28.4%)	2.67
Label makes smoker more likely to quit (mean)	2.43	2.45 (0.8%)	2.61 (7.4%)	2.57 (5.8%)	2.50
Label makes other smokers more likely to quit (mean)	2.30	2.74 (19.1%)	2.97 (29.1%)	2.78 (20.9%)	2.70
Label helps prevent children from smoking	2.45	2.57 (4.9%)	3.13 (27.8%)	3.27 (33.5%)	2.79