Comment on FDA’s Tests of Warning Labels for Cigarettes

Docket No. FDA-2010-N-0568

Daniel Romer and Kathleen Hall Jamieson
Annenberg Public Policy Center
University of Pennsylvania

Submitted on January 11, 2011

We acknowledge the helpful comments of various colleagues, including Ellen Peters of Ohio State University, Paul Slovic of Decision Research and the University of Oregon, and Andrew Strasser of the University of Pennsylvania. The assistance of the team at Research Now, that implemented the research described herein, is also gratefully acknowledged.

Address correspondence to D. Romer, APPC, 202 S. 36th St, Philadelphia PA 19104-3806 or to Dromer@asc.upenn.edu.
Comment on FDA's Tests of Warning Labels for Cigarettes

Docket No. FDA-2010-N-0568

Annenberg Public Policy Center
University of Pennsylvania

The Family Smoking Prevention and Tobacco Control Act of 2009 (FSPTC) requires the FDA to regulate the marketing and advertising of tobacco products, a mandate that includes responsibility for supervising and developing graphic warning labels that will cover 50% of the front and back of all cigarette packages sold in the U.S. Although the U.S. pioneered the use of warnings on cigarette packs and advertising, since 1984 nothing has been done to upgrade or refresh the 4 warning statements placed on the sides of cigarette packs. In the intervening years, other countries, led by Canada, have implemented more graphic visual warning labels. Drawing on the Canadian model, which has been successfully implemented since 2000 (Hammond, Fong, McDonald, et al., 2003; Hammond, Fong, McDonald, et al., 2004), the FSPTC requires that new warnings contain one of nine statements along with a graphic image to draw user attention and to enhance the impact of the warnings.

As a first step in implementing the new warning label system, the FDA used an online survey panel of smokers to test multiple images for each of the 9 warning statements. Each respondent was shown a randomly selected mock-up of a hypothetical cigarette package and asked a battery of questions to assess the effects of the exposure. In control conditions, respondents were simply shown the warning statement on the side of the pack with no image. The results of this study were released early in December 2010 with a request for public comment on the suitability and acceptance of the proposed warnings (see http://www.regulations.gov/#!documentDetail;D=FDA-2010-N-0568-0008).

In order to highlight limitations in that research and identify ways to increase the impact of warning labels, in this comment, we report Annenberg Public Policy Center (APPC) research replicating and extending the study done by the FDA. In particular, we focus on three concerns that make the FDA research less useful than it could have been: (1) the failure to assess the critical effects that mediate the efficacy of warnings, (2) the lack of attention to warning components that could enhance a warning’s efficacy, and (3) the lack of analysis of warning effects on population subgroups that might respond more or less favorably to any warning.

**Failure to Assess Critical Effects**

The APPC in collaboration with researchers at Decision Research has previously tested the effects of exposure to Canadian warning labels with U. S. smokers and nonsmokers (Peters, Romer, Slovic, et al., 2007). In this research, we showed that the graphic Canadian visual-verbal labels attach more negative affect to the smoking experience than the simple print warning statements in place in the U.S. (see warnings 5 and 10 in the Appendix for examples of the Canadian statements and images). While this research only tested short-term effects of the
warnings, theory and research predict that repeated exposure to such messages should increase intentions to quit.

This prediction is based on findings indicating that affect is a powerful cue guiding people’s behavior (Slovic, 2001). More formally known as the affect heuristic (Slovic, 2001), this theory predicts that attaching negative affect to risk information about a behavior (e.g., smoking) reduces that behavior. Survey research with large numbers of smokers and nonsmokers supports the conclusion that risk information associated with negative affect toward smoking is linked with lower levels of smoking, especially among young people (Romer, & Jamieson, 2001). Furthermore, negative affect toward smoking is associated with both increased intentions to quit the habit among smokers and far less trial of smoking among nonsmokers (Romer, & Jamieson, 2001). Indeed, risk perceptions that are not associated with negative affect toward the use of cigarettes are largely ineffective in influencing smoking.

Contrary to our earlier research, the FDA study appeared to offer only weak support for the effectiveness of graphic warning labels. Intentions to try to quit smoking in the next 30 days were rarely enhanced by these warnings (in comparison to control warnings). Although many of the warnings were found to elicit more negative emotional reactions than controls, these effects did not appear to transfer to intentions to quit smoking.

It is noteworthy that the FDA study did not assess the effects of different warnings on the affect attached to the smoking experience. As we noted above, such measures have been used in the past to assess the affect attached to smoking (Romer, & Jamieson, 2001) and the effects of different warning labels (Peters, Romer, Slovic, et al., 2007). To more adequately address this limitation, our research included a measure of this important mediator.

Our study sought to determine whether better measures of the affect heuristic would detect more evidence of negative affect attached to the smoking experience, the hypothesized critical mediator of the effects of risk information.

Importance of Warning Label Components

In addition to the failure to assess affect, the FDA study did not isolate components of the warning stimuli that might contribute to their effectiveness. This approach has been used to study specific components of warnings for alcohol (e.g., Laughery, Young, Vaubel, & Brelsford, 1993). Our testing explored two components: use of color and use of language to contextualize and expand on visuals.

Color. The importance of such analyses can be seen in the warnings tested by the FDA. For example, to illustrate the danger of tobacco use for contracting cancer, the FDA tested a warning label stating, “Cigarettes cause throat, lung, mouth cancer” (see warning 6 in Appendix). The words “throat, lung, mouth” were shown in bright red with a burning cigarette nearby in the same bright color against a black background. This warning performed relatively well in the test despite the fact that the label did not show images illustrating the effects of contracting cancer. We hypothesized that in that label, the use of the color red signaled the presence of danger (cf.
Elliot, & Maier, 2007; Laughery, et al., 1993), an effect that could create greater negative affect toward smoking than the use of a more neutral color even with the same warning content.

The FDA tested an alternative version of this image with greater use of white, but it was not clear from the FDA report whether white replaced the background or the red color of the image. Nevertheless, the white version was less effective despite enabling greater recall of the warning message. This finding was interesting because the warning text actually contained additional information not contained in the text mandated by the FSPTC (that cigarette smoke not only causes cancer in general, but throat, lung, and mouth cancer in particular).

Since the actual alternative version tested by the FDA was not available, it was possible that the greater use of white was restricted to the background rather than the wording and cigarette image, all of which could simply have made the text stand out more. We therefore sought to determine whether a warning component, such as the color red against a black background, could be responsible for the effects of that label. To test this hypothesis, we compared it to the same image replaced with orange coloring, a warning color with less emotional impact (Elliot, & Maier, 2007; Laughery, et al., 1993). If we found that the red warning was more effective than the orange warning, it would further illustrate the importance of testing for such design elements in research to identify more effective warning labels.

**Verbally enhancing the visuals.** Closely associated with our hypothesis regarding color was our observation that the warnings proposed by the FDA did not exploit the full range of warning content that the Canadian labels employ. In particular, the Canadian labels not only show an image with a standard warning statement, but also include additional information to enhance the warning. The red warning used by the FDA was an exception to this practice because it elaborated on the kinds of cancers that are caused by cigarette use. We hypothesized that warning labels with such additional risk information could be more effective than the more general warnings mandated by the FSPTC. For example, in addition to simply saying, “cigarettes are addictive,” the Canadian labels say, “Studies have shown that tobacco can be harder to quit than heroin or cocaine.” Such information explains the scientific basis for the warning and why a cigarette might be akin to injecting a drug into one’s arm (see warning 1 in the Appendix). We therefore tested some of the FDA images with and without such information to further illustrate how warnings can be made more impactful and how experimental variations can be helpful to isolate effective warning components.

**Importance of Population Subgroups**

Our final goal was to examine more carefully the effects of warnings on smokers by age, gender, and other demographic characteristics. The FDA study only reported differences by age with samples in three age ranges: 13 to 17, 18 to 24, and 25 and older. However, it is also important to examine effects of warnings by gender and such characteristics as presence of children in the home. Although a given warning may not appear to be effective for an age group as a whole, it may succeed for certain subgroups. For example, the warning that smoking during pregnancy can harm one’s baby is potentially more effective with younger females than with older adults. This sort of subgroup analysis is important because the success of a warning system will depend
on the overall impact of the entire portfolio across the widely varying smoking and non-smoking populations.

Methods

To ensure that our findings could be compared to those of the FDA, APPC commissioned the same online survey firm used by the FDA, Research Now, to recruit our sample. The criteria for inclusion were identical with the exception that respondents who had participated in the FDA study were not eligible for the study. Because Research Now has a panel of over 3 million participants, we were able to complete the study in a timely manner. The standards of the University of Pennsylvania Institutional Review Board dictated that we differ from the FDA study in one respect; we did not employ the teen panel that was used by the FDA because we were unable to secure approval from parents of those panelists. We therefore restricted our replication to two age groups of current smokers studied in the FDA research, young adults ages 18 to 24 and adults ages 25 and older.

We used many of the same survey questions and procedures employed in the FDA study, but added questions to test the hypotheses articulated above. We limited our test to four of the warning statements mandated by the FSPTC: Cigarettes are addictive, Smoking during pregnancy can harm you baby, Cigarettes cause cancer, and Smoking can kill you. In each case, we included one of the same graphic warnings tested by the FDA. We selected these primarily to illustrate the hypotheses we sought to test. They were not selected because they performed the best in the FDA test. However, for three of them, we included multiple variations that enabled us to identify warning components that might make the images more effective. We also tested two Canadian warning labels that we found to be effective in our previous research.

For each warning, we compared test images against a control warning that only showed the statement on the side of the pack. We tested a total of 16 different warning images, including 12 test images and 4 controls. The entire set of warnings is contained in the Appendix.

Respondents were randomly assigned to one warning condition in which they were asked to view the package in anticipation of answering questions about it later. There were at least 150 respondents in each condition with a somewhat larger number in the older age group. We focus our analysis on the three issues described in the introduction; however, more complete analyses of the results will be reported in subsequent papers.

Findings

Sample Characteristics

Table 1 shows the major demographic and smoking characteristics of the sample. Given that respondents were randomly assigned to warning label conditions, the distribution of the characteristics was not associated with condition. The samples did differ however by age. The older sample had a higher percentage of females than the younger sample (53% vs. 35%), and it had a higher percentage of daily (vs. occasional) smokers (77% vs. 46%). The older sample also had a higher percentage of non-Hispanic white smokers (82% vs. 67%). However, both samples
had about the same percentage who claimed to intend to quit smoking in the next 30 days (about 40%). Although the older sample was more likely to be married, there was no difference in the proportion in either sample that had at least one child under the age of 5 in the household (about 11%).

Table 1. Percentages of respondents in each sample and in total with various demographic and smoker characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Under Age 25 (N=2500)</th>
<th>Over Age 25 (N=2806)</th>
<th>Total (N=5303)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>53</td>
<td>35</td>
<td>44</td>
</tr>
<tr>
<td>Female</td>
<td>47</td>
<td>65</td>
<td>56</td>
</tr>
<tr>
<td>Currently Smoke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Every Day</td>
<td>46</td>
<td>77</td>
<td>62</td>
</tr>
<tr>
<td>Some Days</td>
<td>54</td>
<td>23</td>
<td>38</td>
</tr>
<tr>
<td>Intend to quit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>42</td>
<td>38</td>
<td>40</td>
</tr>
<tr>
<td>No</td>
<td>58</td>
<td>62</td>
<td>60</td>
</tr>
<tr>
<td>Racial-Ethnic Identity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non Hispanic White</td>
<td>67</td>
<td>82</td>
<td>75</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>3.4</td>
<td>6.5</td>
<td>5.1</td>
</tr>
<tr>
<td>Hispanic</td>
<td>17</td>
<td>6.7</td>
<td>12</td>
</tr>
<tr>
<td>Other</td>
<td>13</td>
<td>4.8</td>
<td>7.9</td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10</td>
<td>49</td>
<td>31</td>
</tr>
<tr>
<td>No</td>
<td>90</td>
<td>51</td>
<td>69</td>
</tr>
<tr>
<td>Children under age 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>11</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>No</td>
<td>89</td>
<td>90</td>
<td>89</td>
</tr>
</tbody>
</table>
Are Graphic Visual Warnings More Effective than Warning Statements Alone?

We found evidence across the 12 test warnings that visual warnings on the front of the pack are more effective than simple statements on the side of the pack. To assess the efficacy of the warnings, we used two indices. One was the response to a question about the perceived effectiveness of the package to “discourage someone from smoking.” Possible responses ranged from 1 (not at all) to 4 (very much). This is a standard measure to assess the perceived strength of a media message. Respondents often do not acknowledge the effects of a message on their own behavior, but are quite willing to project such effects on to others. A second and more theoretically critical measure assessed the affect attached to the smoking experience. To assess this, respondents were asked to imagine themselves “smoking a cigarette right now,” and then to report “how good or bad you would feel smoking a cigarette right now.” (Romer & Jamieson, 2001). Possible responses ranged from (1) very good to (4) very bad.

As seen in Figure 1, for all warnings, the graphic labels were more likely to be seen as discouraging others from smoking than the simple warning statement. This effect was weakest for the image of the man in the coffin. However, the warning statement that “smoking can kill you” was the strongest of the four on its own, making it more difficult for the visual to show an effect.

Figure 2 shows the effects of the graphic labels on affective responses to smoking. Here again we see wide effects of the more graphic warnings. With the exception of the man in the coffin and the orange variation for the cancer warning, all of the graphic warnings increased negative feelings toward smoking compared to the control ($p < .05$). The figure shows the proportion of respondents who reported feeling either bad or very bad while smoking. For some of the warnings, this proportion approached 60% of respondents compared to only about 40% for the statements alone.

Similar to the results of the FDA study, we found that virtually none of the graphic warnings produced increased predictions of trying to quit smoking in the next 30 days. However, negative affect toward smoking was consistently related to this prediction across all respondents ($r = .30$, $p < .001$). Furthermore, negative affect was inversely related to agreement with the statement that “I want a cigarette right now” ($r = -.35$, $p < .001$). Thus, the expectation that negative affect toward smoking would be a motivator of reduced smoking was clearly supported. However, a more long-term study of the effects of repeated exposure to graphic warning labels would provide a more appropriate test of this prediction. Indeed, it is overly optimistic to expect that a brief single exposure to a warning label would be sufficient to reduce long-term smoking intentions or behavior.

Do Variations in Warning Label Text or Presentational Elements Matter?

Some of our warning labels were created to isolate differences between warning components. In particular, we wanted to know whether use of the color red to accentuate warning content would enhance the impact of the label. As Figure 2B indicates, the red version of this label did produce a statistically reliable increase in negative affect compared to the control ($p = .011$), while the orange warning was only marginally significant ($p = .06$). Since both of these warnings contained
the same additional language specifying that cigarettes cause throat, lung, and mouth cancer, we can be more confident that the increased effectiveness of the red version is attributable to the use of red in the text and accompanying image of a burning cigarette. Indeed, in post-viewing tests of recognition of the warning statement, there was no difference in recognition of the warning statement between these two conditions. We also found differential effects of this warning on respondents of different age, a finding we discuss in the next section.
Figure 1. Mean rating of how much warning would discourage others from smoking on a 1 (not at all) to 4 (very much) point scale. (A) Addiction series; (B) Cancer series; (C) Harms baby series; (D) Kills series.
Figure 2. Proportions of smokers saying they would feel bad smoking a cigarette right now. (A) Addiction series; (B) Cancer series; (C) Harms babies series; (D) Kills series.
We also tested versions of the same FDA-mandated warning but with enhancements to the information describing the risk. In the addiction series, we added the Canadian text to the FDA cartoon of the cigarette pictured metaphorically as a needle injecting a drug (see warning 2 in Appendix). We found a marginally enhanced effect on affect toward smoking in this condition compared to the other two warning conditions ($p = .066$). This finding suggests that adding information about the risk of addiction could help smokers to appreciate the warning.

The power of the enhanced messages was most apparent in the series designed to carry the message that “smoking during pregnancy can harm children.” We compared the FDA tested image of a cartoon baby in intensive care (warning 13 in Appendix) to two alternatives each with the Canadian statement, “Tobacco use during pregnancy increases the risk of preterm birth. Babies born preterm are at an increased risk of infant death, illness, and disability.” In one version (warning 11 in Appendix), we used the Canadian image of a baby in intensive care. In the other, we used the FDA cartoon image (warning 12 in Appendix). Unfortunately, we were unable to reproduce the Canadian baby image in its entirety, potentially limiting its effectiveness. Nevertheless, both produced strong effects on affect ($p < .001$), and the FDA version with the Canadian message was stronger than the FDA version alone ($p = .024$).

It also noteworthy that a different Canadian version of this message was also quite effective despite only showing the picture of a well-dressed pregnant woman smoking (warning 10 in Appendix). This label included the enhanced message that “Tobacco use during pregnancy reduces the growth of babies during pregnancy. These smaller babies may not catch up in growth after birth and the risks of infant illness, disability and death are increased.” This label was not statistically different from the enhanced FDA image (warning 12) in eliciting negative affect toward smoking ($p = .322$). Thus, language can carry an emotional punch when combined with a picture of someone displaying the behavior.

We also tested a Canadian warning in the cancer series that had done well in our previous research (picture of yellow teeth and diseased gums in warning 5). This label, which included the message that “Cigarette smoke causes oral cancer, gum disease and tooth loss” produced the strongest negative affect toward smoking of all the ones we tested. Although we do not have direct evidence for the separate effects of the image versus the text, it is plausible that the text complemented the image by emphasizing smoking’s multiple effects. Simply saying that smoking causes cancer would not highlight the adverse effects of smoking on the appearance and health of teeth.

**Do Population Subgroups Respond Differently to the Same Warning?**

The study design used by the FDA appropriately segmented the audience by age. However, gender differences can also be important. In the case of messages concerning the harmful effects on children, one would expect that households with children would respond more strongly. We found evidence for all three subgroup effects of gender, age and the presence of children.
As already noted, the use of red coloring to increase the salience and danger associated with smoking was more effective than use of orange coloring with little else changed. However, an analysis of this warning by age and gender indicated that the message was particularly effective in creating an emotional response among older smokers \( (p = .007) \) with a tendency for a stronger effect among women.

We also expected age and gender effects for the message detailing the harmful effects of smoking while pregnant. However, we were surprised to see that the primary differences were related to respondent age and whether the household had children under the age of 5. In general, younger (under age 25) smokers responded more emotionally to this message \( (p = .014) \) than did older ones (over age 25); males did not react less emotionally than females \( (p = .12) \). In addition, households with children under 5 were especially likely to respond with more negative emotion to the experience of smoking \( (p = .006) \). The strongest age difference occurred for the FDA image combined with the Canadian message (warning 12). In this case, the warning was particularly effective for younger respondents irrespective of gender.

We did not observe strong gender or age effects for any of the addiction warnings we tested. However, we did find that the message that smoking can kill was more effective for females \( (p = .007) \). This was true whether the message was merely presented on the side of the pack or with an image of a dead young male.

**Summary and Conclusions**

Our findings indicate that in three of the four warning messages that were mandated by the FSPTC, there was clear evidence that a single exposure to a large graphic warning was more effective in creating immediate negative emotional associations with the act of smoking than exposure to only the verbal warning message on the side of the cigarette package. While we would expect this response to be magnified by repeated exposure to warnings, it did not translate into immediate plans to quit smoking, the primary dependent variable tested by the FDA. However, the strategy behind the use of graphic warnings, at least for current smokers, is to counteract the cravings and other cues that encourage continued smoking. This is hypothesized to occur by attaching negative affect to the experience of smoking over time making it more likely that quit intentions and attempts will be successful. The strategy for those who do not currently smoke would attach negative affect to the experience before addiction sets in. A brief test following a single exposure is unlikely to detect either of these effects.

In the previous study conducted by APPC, both smokers and nonsmokers were shown the entire portfolio of Canadian warnings and their reaction to smoking cues was compared to others who were experimentally exposed to only the current U. S. warning statements. In that instance, we were able to observe the process of negative affect becoming attached to smoking cues. Nevertheless, the best test of the effects of warning labels is to expose smokers to a portfolio of warnings on cigarette packs as they consume cigarettes over time. This form of testing has not to our knowledge been done in any country that has instituted graphic warning labels.
Tests of portfolios of warnings should produce better evidence of the effects of warnings acting both in combination and over time. In this research, we found that some warnings perform better only for some subgroups. If they were carefully selected across an entire portfolio, it might make the portfolio more effective in total than it would be if labels were selected based solely on evidence that they work moderately well across subgroups.

We also found that preliminary tests, such as those conducted by the FDA, could benefit from more careful delineation of warning components to determine the most effective combination. Based on our preliminary tests, we would expect that messages that elaborate on the statements mandated by the FSPTC could increase the effectiveness of warnings. This strategy can educate about the hazards of smoking in greater detail than afforded by the simple statements mandated by Congress. It should also be possible to identify graphic elements, such as the use of color, to enhance the effectiveness of warnings.

The prevalence of smoking remains stubbornly high in the U. S., nearly 25% among non-Hispanic white men ages 18 and older (CDC, 2010). Given the importance and potential of graphic warnings on tobacco products and advertising to reduce tobacco addiction, it will be critical for the FDA and other health agencies to develop effective research programs to maximize the effectiveness of these messages. We hope this initial attempt to illustrate this point will assist in this endeavor.
References


Appendix: Warning labels tested

*Addiction Series*

1) FDA Label   2) FDA image and Canadian Message   3) FDA image and APPC message
4) Addiction Control
Cancer Series

5) Canada Label

6) FDA Label

7) FDA Label Orange Version
Smoking during pregnancy series

10) Canada Version

11) Canada Baby Version

12) FDA Image/Canada Message
13) FDA Label

14) Harm Baby Control
Smoking can kill you series

15) FDA Label

16) Kill you control