Predictive and External Validity of a Pre-Market Study to Determine the Most Effective Pictorial Health Warning Label Content for Cigarette Packages

Li-Ling Huang PhD¹,², James F. Thrasher PhD¹,³, Jessica L. Reid MSc⁴, David Hammond PhD⁵

¹Department of Health Promotion, Education and Behavior, University of South Carolina, Columbia, SC; ²Center for Regulatory Research on Tobacco Communication, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC; ³Department of Tobacco Research, National Institute of Public Health, Cuernavaca, Mexico; ⁴Propel Centre for Population Health Impact, University of Waterloo, Waterloo, ON, Canada; ⁵School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

Corresponding Author: Li-Ling Huang, PhD, Center for Regulatory Research on Tobacco Communication, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599, USA.
Telephone: 919-966-2994; Fax: 919-843-3418; E-mail: huangl@email.unc.edu

Abstract

Introduction: Studies examining cigarette package pictorial health warning label (HWL) content have primarily used designs that do not allow determination of effectiveness after repeated, naturalistic exposure. This research aimed to determine the predictive and external validity of a pre-market evaluation study of pictorial HWLs.

Methods: Data were analyzed from: (1) a pre-market convenience sample of 544 adult smokers who participated in field experiments in Mexico City before pictorial HWL implementation (September 2010); and (2) a post-market population-based representative sample of 1765 adult smokers in the Mexican administration of the International Tobacco Control Policy Evaluation Survey after pictorial HWL implementation. Participants in both samples rated six HWLs that appeared on cigarette packs, and also ranked HWLs with four different themes. Mixed effects models were estimated for each sample to assess ratings of relative effectiveness for the six HWLs, and to assess which HWL themes were ranked as the most effective.

Results: Pre- and post-market data showed similar relative ratings across the six HWLs, with the least and most effective HWLs consistently differentiated from other HWLs. Models predicting rankings of HWL themes in post-market sample indicated: (1) pictorial HWLs were ranked as more effective than text-only HWLs; (2) HWLs with both graphic and “lived experience” content outperformed symbolic content; and, (3) testimonial content significantly outperformed didactic content. Pre-market data showed a similar pattern of results, but with fewer statistically significant findings.

Conclusions: The study suggests well-designed pre-market studies can have predictive and external validity, helping regulators select HWL content.
Introduction
To inform consumers about tobacco-related risks, the World Health Organization Framework Convention on Tobacco Control (FCTC) recommends tobacco product packaging with prominent, rotating health warning labels (HWL) that include pictorial imagery illustrating the consequences of smoking.1 To date, over 60 countries have adopted pictorial HWLs2 but the design of message content varies considerably. To maximize pictorial HWL effects, research is needed on the relative effectiveness of differing HWL content, such as the use of different types of imagery and narrative genres.3 Some experimental studies have addressed these issues but the predictive and external validity of these studies should be determined to help regulators select the most effective HWLs.

Experimental studies indicate that smokers and youth rate graphic, fear-arousing images on HWLs as more effective than other types of HWL imagery.4,5 For example, pre-market experimental studies in Mexico and the United States found that HWLs with graphic imagery that shows gruesome diseased body parts were rated as more effective than both lived experience depicting the social and emotional impact of human suffering from smoking-related disease, and symbolic images using abstract representation of the health effects.3–10 Experiments in Mexico and the United States also indicate that pictorial HWLs with testimonial content focusing on a real personal story work better than HWLs with didactic content,10,11 although this is not always the case, especially among people with higher education.11

While most studies on pictorial HWL content have primarily used experimental, forced-exposure designs with convenience samples, there is limited research on how results from these studies may predict general population responses under conditions of repeated, naturalistic exposure. In order to meet World Health Organization recommendations for rotating multiple HWLs and regular updating of message content, as well as legal challenges initiated by tobacco industry, regulators need robust data on the external validity of pre-market HWL assessments.

The current study examined the predictive and external validity of a pre-market experimental study to assess the efficacy of different pictorial HWL content in Mexico, where the first round of pictorial HWLs were introduced at the end of 2010. Six HWLs were chosen to appear on the Mexican packs by the Ministry of Health based upon findings from focus groups with smoker and nonsmoker adults and youth.12 From 2004 until 2010, HWLs covered 50% of the back of the package and included four text-only messages. Thereafter, pictorial HWLs covered 30% of the package front and text only messages covered 100% of the package back, and two different messages were printed on packages every 3 months. Data for the present study were collected before pictorial HWL implementation, and after smokers had been exposed to six different pictorial HWLs implemented on cigarette packs.

Methods
Sample and Recruitment
Data were analyzed from two sources: (1) a pre-market convenience sample of 544 adult smokers aged 19 years or older who participated in field experiments conducted in Mexico City between June 2 and August 7, 2010, before pictorial HWLs were first implemented in September 2010; and (2) a population-based representative sample of 1765 adult smokers aged 18 years or older from seven major Mexican cities who participated in the wave 5 Mexican administration of the International Tobacco Control Policy Evaluation Survey (ITC Mexico) in 2011, approximately 8 to 9 months after pictorial HWL implementation.

Participants in the pre-market sample were selected using standard intercept techniques of approaching every nth person encountered and inviting eligible adult smokers (who had smoked at least one cigarette in the last month) to participate in the study on the spot.4 A 20-minute computer assisted, interviewer-administered face-to-face survey was conducted using laptop computers. Recruitment sites included two large public parks, a bus terminal, and outside five Walmart stores in Mexico City. Participants in the post-market sample were randomly selected from seven cities (Mexico City, Monterrey, Guadalajara, Puebla, Tijuana, Mérida, and Léon) using a stratified, multistage sampling strategy with face-to-face interviews.13 Within the urban limits of seven cities, census tracts and then block groups were selected with likelihood of selection proportional to the number of households according to census data. Households were selected in random order and visited up for four times to enumerate household members and recruit eligible adult smokers (who had smoked at least 100 cigarettes in their lifetime and smoked at least once a week). The interviewer-administered pencil and paper survey took an average of 45 minutes to complete.

Procedures
HWL Health Effect and Theme
Participants viewed and rated subsets of HWL images, with each set unified by the health outcome of interest (eg, heart attacks, emphysema).9 Each set of HWLs that addressed the same health effect included 5 to 7 warnings, with various themes. Images were drawn by authors from actual HWL implemented in several countries and adapted where necessary. These themes included a didactic text-only HWL, and multiple pictorial HWLs with the same didactic text, but with diverse imagery themes (ie, graphic health effects, lived experience, symbolic images); most HWL sets also included a pictorial HWL with testimonial text, and the same pictorial content as one of the lived experience warnings with didactic text. The theme of each pictorial warning was coded by two independent raters, with conflict resolved by a third rater before the study. The themes were defined as follows: (1) Graphic health effects: vivid depiction of physical effects; (2) Lived experience: depiction of personal experience, including social and emotional impact, or implications for quality of life; (3) Symbolic images: representation of message using abstract imagery or symbol; and (4) Testimonial: warnings with a brief narrative describing a personal consequence of smoking, written as a quote from a person in the image, accompanied by their name and age.9

Rating and Ranking Tasks
In the pre-market sample, participants were randomized to view 2 of 17 sets of HWLs on a computer screen (see all the HWLs in Hammond et al.4). HWLs within each set were presented in a random order and participants rated each one while the image appeared on screen. The six HWLs that later appeared on Mexican cigarette packs were included in six of the 17 health effect sets (ie, toxicity, gangrene, emphysema, secondhand smoke, death, fetal harm; Table 1). After rating each HWL in a set, participants were shown the full set of HWLs (with onscreen position randomized) and asked to rank order them according to which would be most effective for discouraging smoking.
Participants in the post-market survey were shown laminated cards with images of each of the six HWLs that had appeared on cigarette packs by the date of the interview (in random order), and if they recalled having seen it on cigarette packs, were asked to rate that warning. After rating each of the six HWLs, participants were shown one of seven health effect sets (ie, addiction, aging, impotence, lung cancer, stroke, throat cancer, quitting) that included the same stimuli as the pre-market survey and were then asked to rank order each HWL within the set in the same manner as the pre-market survey. The seven health effect sets were selected because they had more balanced themes compared to other sets. To prevent ordering effects, the arrangement of the HWLs on the laminated card was counterbalanced across interviewers.

Measures

HWL Responses

Ratings
Pre-market sample participants rated each HWL on a scale of 1 to 10 (1 = “not at all”; 10 = “extremely”) for whether the HWL: grabbed their attention; was believable; was frightening; would make smokers want to quit; and, overall, how effective they considered the HWL. Effectiveness scores ranged from 1 to 10 (the original scale was 1 to 5 for the post-market sample and was rescaled to 1 to 10). Superscript letters denote significant difference at \( P < .05 \) for all pair wise comparisons of HWL effectiveness scores within the same sample set (a, b, c, d for pre-market sample set; e, f, g for the post-market sample set). HWLs with the same superscript letter are not significantly different from one another. Significance remains the same after adjusting for, sex, age, educational attainment, smoking intensity, and quit intentions.

The authors would like to thank photographer, Pablo Herrerias, for taking the original photos of the cigarette packs in Table 1 and the research team for adapting warning images from other countries.

<table>
<thead>
<tr>
<th>Health effect</th>
<th>Theme</th>
<th>English translation</th>
<th>Pre-market sample</th>
<th>Post-market sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>n=62</td>
<td>n=700</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Toxicity</td>
<td>4.15\textsuperscript{a}</td>
<td>7.11\textsuperscript{f}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smoking make you lose more than one member</td>
<td>5.87\textsuperscript{b}</td>
<td>7.17\textsuperscript{e}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stop smoking now... later will be too late</td>
<td>6.18\textsuperscript{c}</td>
<td>7.61\textsuperscript{f}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smoking kills you... ...and not just you</td>
<td>6.84\textsuperscript{d}</td>
<td>7.55\textsuperscript{f}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smoking does not hurt only you</td>
<td>6.93\textsuperscript{d}</td>
<td>7.62\textsuperscript{f}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>With each puff... ...you can lose it</td>
<td>6.99\textsuperscript{d}</td>
<td>7.65\textsuperscript{f}</td>
</tr>
</tbody>
</table>


Table 1. Effectiveness Scores of the Six Health Warning Labels (HWLs) on Mexican Cigarette Packs

Themes: S, symbolic; LE, lived experience; G, graphic. Items used to form the Effectiveness scale showed good internal consistency (\( \alpha = 0.79–0.90 \)), done for each HWL across samples; five items were whether the HWL: grabbed their attention; was believable; was frightening; would make smokers want to quit; and, overall, how effective they considered the HWL. Effectiveness scores ranged from 1 to 10 (the original scale was 1 to 5 for the post-market sample and was rescaled to 1 to 10). HWLs with the same superscript letter are not significantly different from one another. Significance remains the same after adjusting for, sex, age, educational attainment, smoking intensity, and quit intentions.

Analysis

Unadjusted and adjusted linear mixed effects models were used to test the relative ratings of effectiveness for the six HWLs that appeared on cigarette packs. Unadjusted and adjusted logistic mixed effects models were estimated in which effectiveness rankings served as outcome, and examined which HWL theme was most effective in the six health effects sets that had maximal variation in themes (Supplementary Appendix 1). HWLs that exhibited more than one characteristics (eg, graphic and symbolic in themes), were coded as having both characteristics. Mixed-effect models accounted for repeated measures within individuals, and all models adjusted for sociodemographics (ie, sex age, educational attainment), smoking intensity, and quit intentions. Models for theme analyses were also adjusted for health effect set viewed.

Results

Sample Characteristics

Compared to the analytic sample for the post-market study (n = 1529 for ratings analysis; n = 1427 for ranking analysis; see Supplementary Appendix 2), the analytic sample for the pre-market study (n = 335 for ratings analysis; n = 309 for ranking analysis) included a lower proportion of: males (ratings = 50% vs. 64%;
Models were adjusted for sex, age, educational attainment, health effect, smoking intensity, and quit intentions.

**H WL Ratings**

Pre- and post-market data showed similar relative ratings across the six HWLs that appeared on cigarette packs, with the least and most effective HWLs consistently differentiated from other HWLs (Table 1). The HWL (Label F) that depicted both graphic and symbolic themes showing a fetus surrounded by cigarette butts was rated as the most effective HWL. In both pre-market and post-market samples, but was significantly different from the next most effective HWL (Label E) using a lived experience theme only in the post-market sample. The HWL (Label A) describing toxic tobacco constituents including a deadly poison used to kill rats, coded as a symbolic theme, was rated as the least effective HWL in both samples, but was significantly different from the next least effective HWL (Label B) with a lived experience theme only in the pre-market sample. The other three lived experience HWLs (Labels C, D, and E) had similar effectiveness scores in both samples. However, these three lived experience HWLs were rated significantly differently from the other HWLs (Labels A, B and F) in the post-market sample, but were not completely discernible from the other HWLs (Labels B and F) in the pre-market sample.

**H WL Ranking of Themes**

Table 2 shows estimates from three models in which ranking as “most effective” within a health effect set served as the outcome and HWL themes as the predictors. The first model, comparing text-only HWLs with pictorial HWLs with the same text, indicated a significantly greater likelihood of pictorial HWLs being ranked as the most effective in both samples. The second set of models compared different pictorial HWL themes, and pre- and post-market results showed a similar pattern: graphic and lived experience content (alone or combined) were more likely to be ranked as more effective than symbolic content. However, this pattern was statistically significant only in the post-market sample. In the pre-market sample, statistical significance was reached only for HWLs with graphic content compared to symbolic content in the unadjusted model, and when graphic and lived experience were combined in both unadjusted and adjusted models. The third set of models compared pictorial HWLs with testimonial text to the HWLs that shared the same imagery but used standard didactic text. The results indicated that testimonial content was ranked as significantly more effective than didactic content for the post-market sample (AOR 2.64, 95% CI = 2.27, 3.08), with no statistically significant result for the pre-market sample.

**Table 2. Relative Effectiveness of Health Warning Label Themes**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Theme</th>
<th>n</th>
<th>Probability</th>
<th>Unadjusted</th>
<th>Adjusted</th>
<th>Odd ratios (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-market</td>
<td>Text-only</td>
<td>67</td>
<td>.18</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pictorial</td>
<td>294</td>
<td>.23</td>
<td>1.33 (.01, 1.78)*</td>
<td>1.33 (.01, 1.78)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symbolic</td>
<td>109</td>
<td>.25</td>
<td>1.38 (.01, 1.98)</td>
<td>1.33 (.01, 1.93)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lived experience</td>
<td>90</td>
<td>.29</td>
<td>1.42 (1.01, 1.99)*</td>
<td>1.36 (0.96, 1.93)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Graphic</td>
<td>126</td>
<td>.3</td>
<td>1.53 (1.05, 2.24)*</td>
<td>1.53 (1.04, 2.23)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Graph with lived</td>
<td>76</td>
<td>.32</td>
<td>1.53 (1.05, 2.24)*</td>
<td>1.53 (1.04, 2.23)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Didactic</td>
<td>185</td>
<td>.51</td>
<td>1.08 (0.81, 1.45)</td>
<td>1.07 (0.80, 1.43)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Testimonial</td>
<td>67</td>
<td>.05</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Post-market</td>
<td>Text-only</td>
<td>1358</td>
<td>.67</td>
<td>7.69 (5.97, 9.90)**</td>
<td>7.84 (6.07, 10.11)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pictorial</td>
<td>439</td>
<td>.27</td>
<td>2.22 (.83, 2.70)**</td>
<td>2.22 (.81, 2.72)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symbolic</td>
<td>400</td>
<td>.31</td>
<td>2.42 (2.02, 2.91)**</td>
<td>2.44 (2.02, 2.95)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lived experience</td>
<td>627</td>
<td>.33</td>
<td>2.32 (1.85, 2.91)**</td>
<td>2.29 (1.81, 2.90)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Graphic</td>
<td>206</td>
<td>.32</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Didactic</td>
<td>538</td>
<td>.38</td>
<td>2.68 (2.30, 3.12)**</td>
<td>2.64 (2.27, 3.08)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Testimonial</td>
<td>881</td>
<td>.62</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

CI = confidence interval.

*Probability of being ranked as most effective.

*Models were adjusted for sex, age, educational attainment, health effect, smoking intensity, and quit intentions.

Significance levels: *P < .05; **P < .01; ***P < .001.
to pictorial HWLs more than didactic contents. Testimonial content may be more effective than didactic content because it enhances emotional arousal. HELs that employed symbolic representations were perceived as less effective than other pictorial themes, perhaps because abstract depictions are less well understood or because they elicit weaker emotional responses. Furthermore, certain health effects may be more amenable to depiction through particular themes (eg, symbolic images illustrating addiction or quitting; graphic and lived experience images illustrating cancer and disease), as with antismoking advertisements. While it appears that HWL imagery that combines graphic imagery and lived experiences is most powerful, more research is needed on how best to communicate about other health topics, such as cessation advice.

Limitations of this study include the modest sample size for the pre-market study, variations in range and item wording of rating scales between the pre-market and post-market studies, and timing issues of the pre-market and post-market studies. For example, the difference in the ranges of the scales used in the studies (ie, rescaling from 1–5 to 1–10) may have resulted in a narrower range in the post-market ratings, making direct comparisons in ratings between the two samples challenging. The timing of the pre-market and post-market studies could have affected our outcome measures and analytic sample size for the six HWL on the Mexican cigarette packs. It is possible that the timing of pre-market study 3 months before HWL implementation meant that some participants had previously viewed the HWLs because of media coverage of the new HWLs. However, it seems unlikely that a study conducted considerably earlier than policy implementation would produce different results. In the period before implementation of pictorial HWLs, which were just one part of a comprehensive tobacco control law, media attention to tobacco was focused more on economic issues and smoke-free policies. There was no coordinated media outreach during the runup to implementation, and informal monitoring of media suggested that coverage remained null to low during the study period. Furthermore, the implementation date was for printing of the HWLs, not for their appearance in stores, which can take up to 6 months from the print date. Hence, we do not think we would have found different results if the pre-market study had been conducted earlier. The substantially different sample size of post-market study for each HWL was likely affected by the timing of the six HWL introduced in Mexico and the post-market assessment. For example, one HWL (Label F) was recalled by many more participants than other HWLs. This HWL was one of the first two to appear on market, so it was more likely to be recalled due to longer period of exposure. Furthermore, future validation research in this area should involve more comparable samples and ensure adequate power for examining key hypotheses. Nevertheless, this research represents a unique advance due to sharing of protocols across pre- and post-implementation samples, including a population-based representative sample of smokers in the post-market survey.

In conclusion, this study suggests that well-designed pre-market studies can have predictive and external validity, and may be used to assist regulators in making evidence-based decisions about the selection of HWL content. The pre- and post-implementation design provides a model to systematically test and evaluate the most effective message themes for the general public or particular target groups.

Supplementary Material
Supplementary Appendices 1 and 2 can be found online at http://www.ntr.oxfordjournals.org

Funding
Data collection and analyses for this project were supported by grants from the US National Cancer Institute (P01 CA138389 and R01 CA167067), the Mexican National Council on Science and Technology (Salud-2007-C01-70032) and also by the Propel Centre for Population Health Impact, a Canadian Institutes of Health Research New Investigator Award, and the Canadian Cancer Society Research Institute Junior Investigator Award (awarded to DH). The funding sources had no role in the study design, in collection, analysis, and interpretation of data, in the writing of the report, or in the decision to submit the article for publication.

Declaration of Interests
None declared.

Acknowledgments
The authors would like to thank Cődice Comunicación Diálogo y ConcienSC and Instituto Nacional de Salud Pública (INSP) for their assistance in conducting this work, and to acknowledge the assistance of Ana Dorantes (field manager for Mexico) and Samantha Daniel.

References


