International Journal of Management and Marketing Research
Vol. 9, No. 1, 2016, pp. 81-96
ISSN: 1931-0269 (print)
ISSN: 2157-0698 (online)

DO CONSUMERS RECALL PRODUCTS' WARNING LABELS? A META-ANALYSIS

Mohamed M. Mostafa, Gulf University for Science and Technology, Kuwait

ABSTRACT

Meta-analysis is a statistical technique that allows one to combine the results from multiple studies to glean inferences on the overall importance of a certain phenomenon. This study employs a substantive meta-analysis approach to quantitatively summarize the results of empirical studies of the direct impact of products' warning labels on consumers' recall. When all the available estimates are combined and averaged, there seems to be a genuine and positive effect of warning labels on consumers' recall (average effect size = 0.34, aggregate n = 1882). The findings of this study significantly refine the body of knowledge concerning the impact of products' warning labels on consumers' recall, and thereby offer an improved conceptual framework for marketers and warning label designers.

JEL: M30, M31, M37

KEYWORDS: Warning Labels, Consumer Recall, Meta-Analysis

INTRODUCTION

The proliferation of warning labels has become a defining characteristic of modern marketplace. Cigarettes, alcoholic beverages, appliances, saccharine, cosmetics and other personal care products are among consumer products that carry warnings. Various service products also carry warning labels. Examples include investment and insurance services. The presence of warning labels is not restricted to product packaging and package inserts; they also appear in the advertising of various products and services and in places where products are sold (Stewart and Martin, 1994). Warning labels serve two major aims. The first aim is consumer awareness, that is, education about potential problems associated with excessive or inappropriate consumption. The second aim is the prevention and/or modification of harmful consumer behaviour ensuing from education (Stockley, 2001). Warning labels have been consistently proposed as a cost-effective prevention strategy to communicate risks associated with certain products. For example, alcoholic warning labels have been shown to receive the highest public support among a variety of alcohol control policies, including policies on availability, educational programmes, and higher taxes (Kaskutas, 1993).

For products that are potentially hazardous to consumers, warning labels are potentially important communication tools that are available for both marketers and public policy agencies. Such warning labels can inform consumers about the risks and potential dangers related to product usage (Lepkowska-White and Parsons, 2001). Persuasion literature indicates that message recall affects consume attitudes towards the message, which in turn may influence subsequent behavior (Ajzen, 1991). Curiously, given the importance of the relationship between warning labels and consumers' recall, no systematic review has been conducted to synthesize the results of previous research. By sifting through previous research we fill this research gap by answering the following question: "Does the exposure to warning labels affect the consumers' recall of their contents?"

Thus, in this research we conduct a meta-analysis to synthesize and validate the impact of warning labels on consumers' recall and memory by analyzing the empirical results of various studies over a period spanning around 25 years. Synthesizing existing research has an impact beyond its utility for informing policy and practice as systematic reviews enable knowledge to be accumulated in a manageable way (Harden and Thomas, 2005). This research is organized as follows: The next section briefly reviews previous literature. The methodology section describes the methodology used to conduct the analysis. The results section provides the results of the quantitative analysis. This is followed by the conclusions section, which concludes the paper by providing the implications of the results and exploring avenues for future research and practice.

LITERATURE REVIEW

Lehto and Miller (1986) argue that the major purpose of warning labels is to alert consumers to a potential hazard and trigger the processing of additional information in memory, within the warning itself or elsewhere. Across several studies, effects of warning labels have been reported on awareness, exposure, and recall (e.g. Hankin et al., 1998). Warning effects on memory and recall are important because they are measures of the effectiveness of the warning. Recall and memory are also important because changes in these cognitive measures may ultimately lead to reductions in product-related problems (MacKinnon et al., 2000). Previous research on recall of warning labels has been sporadic and limited (Morris, Gilpin, Lenos & Hobbs, 2011). A stream of research has focused on warning labels' visibility and consumers' awareness (Fox, Krugman, Fletcher, & Fischer, 1998, Krugman, Fox, Fletcher, Fischer, & Rohs, 1994, Marin, 1997). Extended exposure to products' warning labels has been found to increase both awareness and control behaviors (Muggli, Pollay, Lew, & Joseph, 2002). Warning label wording has also been found to affect the effectiveness of such labels (Heaps & Henley, 1999). For example, the effectiveness of the warning "Smoking causes lung cancer, heart disease, emphysema, and may complicate pregnancy" was rated as more effective compared to the label "Tobacco smoke harms the health of individuals near the smoker" (Malouff, Schutte, Frohardt, Deming, & Mentelli, 1992). Similar results have been reported by Andrews, Netemeyer, & Durvasula (1991), Beltramini (1988), and Loken and Howard-Pitney (1988).

Previous marketing research has also investigated consumers' compliance behavior with warning labels. Mixed results have been reported showing varying rates of behavioral compliance. For example, Orwin, Schucker, & Stokes (1984) found that a saccharin label warning label against elevated cancer risk reduced diet soft drink by around 4 % below predicted levels. In a similar vein, Goyal, Rajan, Essien, and Sansgiry (2012) found that organ-specific warning labels improve consumers' risk perception of liver damage and increase the intention to perform protective behaviors regarding over-the-counter acetaminophen products. Borland et al. (2009) found that cigarettes' warning labels help increase quit-related cognitive responses. Miller, Hill, Quester, & Hiller (2009) also found that warning labels associated with graphic fear appeals result in an increase in the number of calls to quit lines. Hammond, Fong, McDonald, Brown, and Cameron (2004) found that around 45% of the respondents reported either benefiting from warning labels by quiting or by staying abstinent. However, in a qualitative investigation spanning four countries, Haines-Saah, Bel, and Dennis (2015) found that cigarettes' warning labels might have potential unintended consequences on consumers by obscuring the social and embodied contexts of the smoking experience.

Yong et al. (2014) used structural equation modeling (SEM) to investigate the impact of warning labels on smoking behavior. The authors found that warning labels influence thoughts about risk of smoking, increase worry about negative outcomes, and are also strong predictors of subsequent quit attempts. In a stated preference choice experiment, Lacanilao, Cash, and Adamowicz (2011) asked respondents to choose between high-fat snakes, some displaying a warning label, and some healthier snakes. The authors found that one class of respondents heeds warning labels, another class cares only for price but not warning labels, yet a third class avoids less healthy snacks and cares about price only when a warning label is present. However, using eye-tracking techniques, Crespo, Cabestrero, Grizb, and Quiros (2007) found that using graphic images along with warning labels decrease the effectiveness of warning labels compared to using the traditional "general surgeon" warning labels. Ruiter and Kok (2005) found that verbal warning labels were only deemed to be effective when consumers were directed to "engage with the text". This is particularly true when the wear-out effect of warning labels is present.

Studies reporting alcohol warning labels recall found that there was an increase from 3.8% to 28.5% between 1989 and 1994 (Mazis, Morris, & Swasy, 1995). The drink driving text warning was recalled by around 25% of adults in the US (Tam & Greenfield, 2010). However, some studies found that such text warnings had virtually no impact on behavioral change (Scholes-Balog, Heerde, & Hemphill. 2012, Wilkinson & Room, 2009). Some studies found that alcohol drinkers were unable to recall the language of alcohol warnings (MacKinnon et al. 2001, Malouff et al. 1993). Barlow and Wogalter (1993) found that alcohol warnings in print ads were recalled more when the warning was highly conspicuous. Other studies have reported that text-only warnings were forgettable and ineffective compared to graphic pictorial warnings, which create negative effect toward smoking and increase willingness to quit (Bansal-Travers, Hammond, Smith, & Cummings, 2011, Moodie, MacKintosh, & Hammond, 2009, Peters et al. 2007, White, Webster, & Wakefield, 2008). However, Kees et al. (2010) found that gruesome pictorial warnings are less recalled because they interfere with the cognitive processing and comprehension of the warning message text. These findings contradict a recent survey conducted in Mexico, in which the authors reported a positive association between pictorial health warning labels and free recall in a sample of 1765 respondents (Thrasher et al. 2012).

DATA AND METHODOLOGY

Meta-analyses are being used more frequently for setting public policy, defining research agendas, and influencing practice. Meta–analysis is a statistical synthesis method that provides the opportunity to view the “whole picture” in a research context by combining and analyzing the quantitative results of many empirical studies (Glass, 1976). The aim is to take advantage of the large numbers resulting from the use of a number of samples to permit a better estimate of the population statistic. It is likely to prove particularly useful where the samples in individual studies are too small to yield valid conclusions or where there are many non-experimental studies with low statistical power (Petitti, 1994). Rosenthal and DiMatteo (2001, p. 62) argue that meta-analysis is more than a statistical method; it is a “methodology for systematically examining a body of research, carefully formulating hypotheses, conducting an exhaustive search, recording and statistically synthesizing and combining data and reporting results.”

The primary advantage of a meta-analysis is that findings from a multitude of studies can be effectively summarized using a standardized metric of effect (e.g., Cohen’s d) that is relatively easy to interpret. The effect size statistic ranges between positive 1 and negative 1 in the vast majority of cases. Although no standards exist for interpreting effect size estimates, Cohen (1977) suggests that when interpreting values of d, .20 would be indicative of a small effect, .50 a medium effect, and .80 a large effect.

Analogous to most meta-analysis, our study began with an examination of the known research on consumer recall of warning labels. This examination was conducted by searches of the standard computerized databases (e.g., ABI/Inform, EBSCO, Science Direct, Emerald etc.) From the relevant literature, citations and references were extracted and catalogued in an iterative process until nearly 40 articles, proceedings, book chapters, and dissertations were identified that in some fashion addressed the consumer recall of warning labels phenomenon. The first contribution was published in 1982. Articles were distributed across twenty-three journals, proceedings and dissertations. Studies excluded were (a) studies that made references to consumer recall of warning labels but did so in the context of brief conceptual discussion and thus was only tangentially related tour investigation, (b) studies that did employ some empirical analysis, but did not report Pearson correlation coefficients or other statistics capable of being converted to point-biserial correlations, and (c) studies known to be drawn from the same data set as another study that has already been included in the analysis. This resulted in 14 studies in the meta-analysis (Table 1).

Table 1: Summary of Studies Included In Meta-Analysis

Author(S)	Research Design	Product	Location of Warning Label	N	R
Barlow and Wogalter (1991)	Experimental	Alcohol	Adverts	105	0.66
Bhalla and Lactovicka (1984)	Experimental	Cigarettes	Adverts	84	0.44
Desaulniers (1987)	Experimental	Chemicals	Product	50	0.20
Frantz (1992)	Experimental	Chemicals	Product	80	0.49
Gardner-Bonneau et al. (1989)	Experimental	Cigarettes	Paper	118	0.09
Goldhaber and deTurck (1988)	Experimental	Pool	Wall	328	0.28
Kalsher et al. (1991)	Experimental	Alcohol	Poster	134	0.31
Karnes and Leonard (1986)	Experimental	Chemicals	Product	40	0.35
*Nohre et al. (1999)	Survey	Alcohol	Product	6391	0.31
Otsubo, S. (1988)	Experimental	Saw	Product	125	0.20
Strawbridge (1986)	Experimental	Chemicals	Product	195	0.16
Wogalter et al – study 1 (1992)	Experimental	Chemicals	Sign	48	0.51
Wogalter et al – study 2 (1992)	Experimental	Chemicals	Sign	80	0.38
Young and Wogalter (1990)	Experimental	Electric Generator	Paper	531	0.30

**Note: Excluded from meta-analysis. This table shows all the studies included in our meta-analysis along with publication dates, sample size and correlation coefficient either reported or calculated from the studies data. As seen from the table, almost all warning labels fall in one of three major categories: cigarettes, alcohol, or chemical products. Correlation coefficients ranged between 0.09 and 0.66.*

Selected studies were coded for classifying study variables. The code book allowed for the collection of a large mount of information such as general study characteristics, design characteristics, demographic information pertaining to the study participants, psychometric properties of the study instruments, methods of statistical analysis and other characteristics.

Meta-analysis procedures suggested by Hunter and Schmidt (2004) were employed in this study. These methods can be outlined as: (a) calculation of effect sizes pertaining to the relationship between warning labels and consumers' recall, (b) analysis of outliers and missing data, (c) homogeneity analysis of effect sizes, and (d) testing for publication bias.

Data were the correlation coefficients from previous studies that described the relationship between warning labels and consumers' recall. Two approaches, namely a single finding per study or multiple findings per study, were often used in meta-analysis. According to Grewal et al. (1997), the advantage of extracting one finding per study is a guarantee of the independence of each finding. Following Palmatier et al (2006), we accumulated results and calculated the average effect within a study when a study gave separate reports for different findings but used the same subjects. For their relative convenience in analyzing a large number of variables, MetaWin version 2.0 (Rosenberg et al., 2000) and R software packages were used to conduct the analysis.

RESULTS

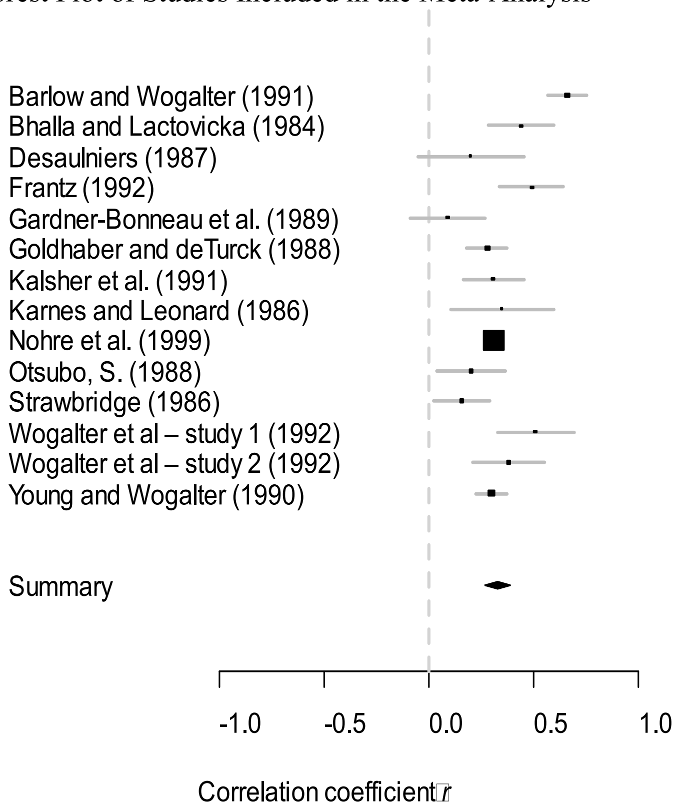
Effect Size

The term effect size refers to the sample estimate of the population effect regardless of the particular effect size indicator used (Fern and Monroe, 1996). Effect sizes were reported in the original papers in a variety of forms (e.g. t, F, P values). We converted these various test statistics to Hedge's d following the formulas suggested by Rosenthal (1991).

In meta-analysis, both fixed-effect models and random-effect models have been used to estimate effect size. fixed-effect models assume that samples within a study share a common true effect size, while random-effects models assume that studies have a common mean effect, but that there is also random variation in effect sizes within a study. Several researchers have used fixed-effect models (e.g. Moller and Thornhill, 1998), although random-effect models are probably more appropriate because there is no reason to assume that there is a single common true effect size across all studies (Gurevitch and Hedges, 1999). We took a random-effects perspective, which Hunter and Schmidt (2000) recommend for routine use in meta-analysis when calculating mean correlations. In this approach, the variability in findings across studies is treated as arising from both the sampling of studies, reflecting between-studies variance, and the sampling of individuals within studies, reflecting sampling error. According to Hunter and Schmidt (2000), the random-effect approach produces more generalizable results and is less subject to Type I errors when the assumptions of the fixed-effect model are not met. I fact, Chen and Peace (2013) argue that in practice most researchers "perform both a fixed-effects and a random-effects meta-analysis on the same set of studies - even if there is an 'a priori' basis for believing the fixed-effects model is appropriate." (p. 68).

The analysis of effect sizes adjusted for measurement error was indicative of a moderate relationship between consumers' exposure to product warning labels and recall (Pearson correlation r = 0.34). 95% Confidence interval for all effect sizes ranged from 0.25 to 0.42 for the random effects model. For comparative purposes, the confidence interval ranged from 0.27 to 0.35 for the fixed effects model. Figure 1 presents a "forest plot" of the estimates and their corresponding 95% confidence intervals for each study included in our meta-analysis, as well as the global summary or combined estimate.

Figure 1. Forest Plot of Studies Included in the Meta-Analysis

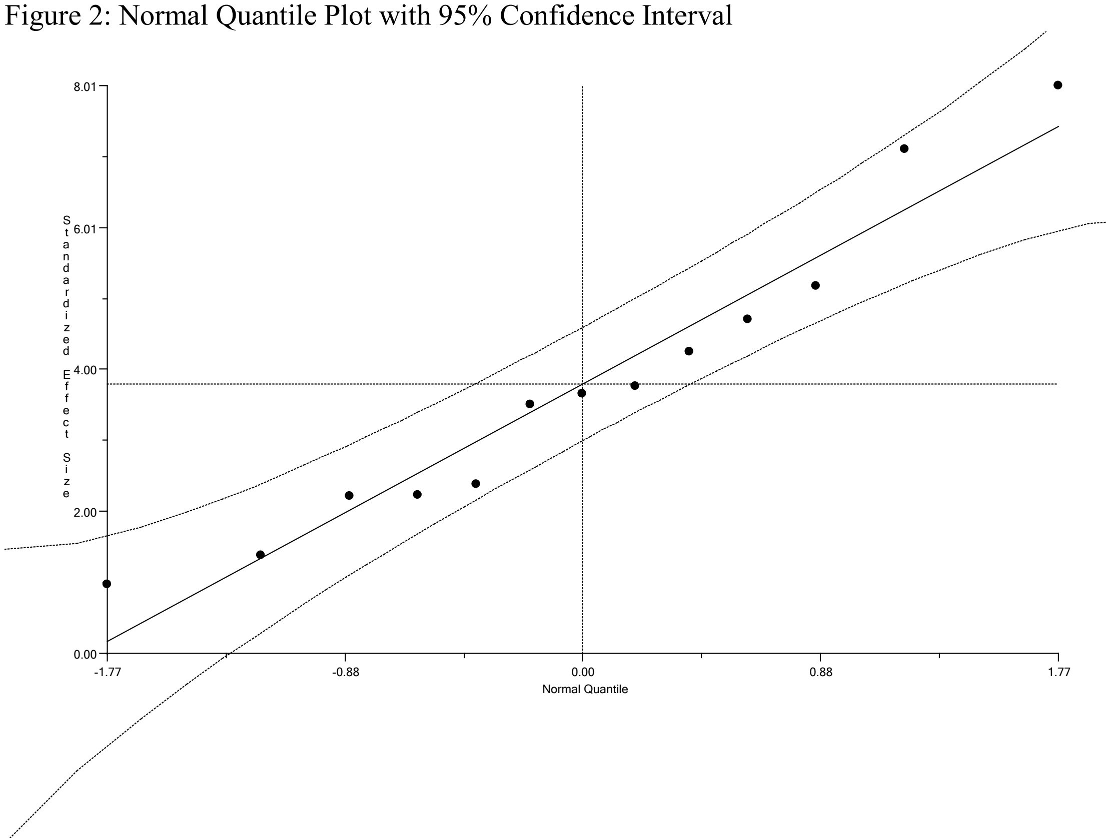


This figure shows a plot of the estimates and their corresponding 95% confidence intervals (CIs) for each study included in the meta-analysis. The figure also shows the global summary of combined estimate. The 95% CIs are represented by lines, while the squares in the middle represent the point estimates. The global estimate is represented by the diamond, whose its width is the associated 95% CI.

Detecting Outliers

Outliers' analysis is an essential step in any meta-analysis study as the presence of outliers is likely to distort the results (Davis and Rothstein, 2006). One simple strategy for handling outliers is to delete them from analysis if they are not believed to be representative of study findings. Analysis of the distribution of the effect sizes revealed one potential outlier that contributed to a skewed effect size distribution. This outlier was reported in Nohre et al.'s study (1999) in which a very big sample was used (n = 6391). After excluding this study, the distribution of standardized effect sizes approached normality as shown in Figure 2.

Figure 2: Normal Quantile Plot with 95% Confidence Interval



This figure shows the normal quantile on the x-axis and the standardized effect size on the y-axis. The normal quantile plot is generally used to detect potential outliers that may contribute to a skewed effect size distribution. After eliminating a potential outlier (Nohr et al. 1999 study), the distribution seems to be fairly normal.

Heterogeneity and Resampling Tests

Assessing heterogeneity means answering the question "How different are the results of these studies?" (Attia et al., 2003, p. 300). The most commonly used test for between-study heterogeneity in meta-analysis is the Q statistic (Fleiss, 1993). The Q statistic is defined as the sum of the weighted squared differences of the effect size of each study from the summary effect size obtained under fixed-effects assumption. Alternative tests have been proposed that usually, but not always, give very similar results (Ioannidis et al., 2006). However, bootstrapping validation suggests that the Q statistic is the most reliable of these tests (Takkouche et al., 1999). We used the Meta package in R to conduct the heterogeneity analysis, which resulted in a significant Q-value (Q = 44.28, df = 12, P = 0.0001), which suggests that the variances in the sample of effect sizes were not homogeneously distributed and could not be accounted for by sampling error alone.

The parametric model used in our study to derive the effect size d relies on the assumption that the observations are normally distributed for each study. In addition, the Q statistic used to assess the homogeneity of the effect sizes among studies is approximately chi-squared distributed when the normality condition is met. If this condition is violated, the conventional tests of homogeneity may be flawed (Hedges and Olkin, 1985).

Adams et al. (1997) recommend that resampling methods should be incorporated in meta-analysis studies to ascertain the robustness of estimates. resampling methods, preeminently the bootstrap technique, test the significance of a statistic by generating a distribution of that statistic by permuting the data many times, each time recalculating the statistic. By comparing the original statistic to this generated distribution, a significance level can be determined (Manly, 1991).

We first computed the bootstrap confidence limits for the mean effect sizes using the conventional method, the percentile bootstrap (Efron, 1979). One problem with percentile bootstrap confidence limits however, is that as sample sizes decrease, the lengths of percentile bootstrap confidence limits tend to become underestimated (Efron, 1987). Bias-corrected percentile bootstrap confidence limits have been suggested to correct for distributions when > 50% of the bootstrap replicates are larger or smaller than the observed value, which happens often with small samples (Efron, 1979). As our meta-analysis contained a small number of studies, we also calculated the Bias-corrected percentile bootstrap confidence limits. As recommended by Adams and Anthony (1996), 5000 replications were done to reduce variation around the significance levels. Results show that the percentile bootstrap confidence limits are quite similar to the parametric confidence limits, which ascertain the robustness of standard meta-analytic techniques.

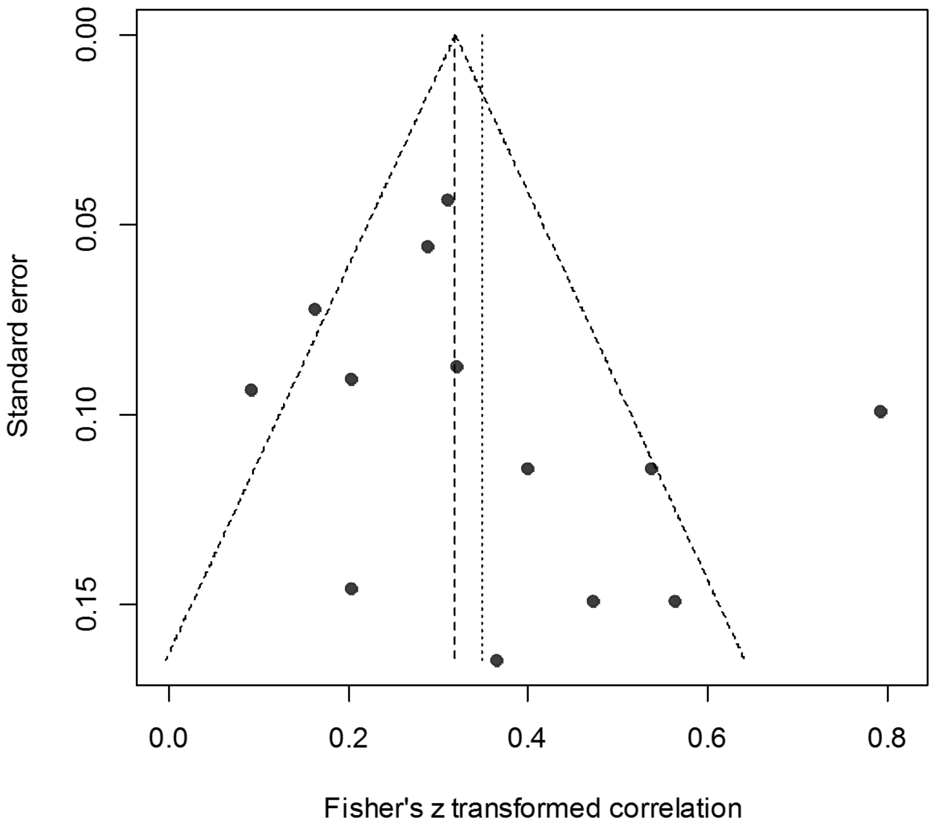
Publication Bias

Publication bias refers to the tendency to publish studies that show statistically significant results. Such "positive" studies are "therefore more likely to be located for and included in meta-analyses, which may introduce bias." (Sterne et al., 2000, p. 1119). A survey of studies reported in four psychological journal and three medical journals found high proportions of papers confirmed the experimental hypotheses being tested (Sterling et al., 1995). Gregoire et al. (1995) reported that 78% of identified meta-analyses of randomised trials had language of publication restrictions. Because of publication bias, studies included in a meta-analysis may not generally represent a random sample of all studies actually conducted (Hsu, 2002). Since all but three of the studies used in this meta-analysis were from published sources in English language, there is a possibility that publication bias is affecting results.

Several methods have been proposed to detect the existence of publication bias in meta-analyses. We used three methods for assessing the potential importance of publication bias. First, we used the funnel plot recommended as a graphical device for investigating the possibility of publication bias (Light and Pillemer, 1984). A common interpretation is that a symmetric, inverted funnel shape implies no publication bias, but if the funnel appears to be missing points in the lower corner of the plot, there is potential bias (Borenstein, 2005).

In Figure 3 a funnel plot was constructed with standardized correlations on the x-axis and standard error on the y-axis. As seen in Figure 3, the funnel plot appears to be relatively symmetric, although it is clearly not perfectly so, indicating little or no publication bias. However, Tang and Liu (2000) criticized the simplistic conclusion that a meta-analysis is biased based on the informal observation of an asymmetric funnel plot and suggested that one should interpret it as a 'precision-related' heterogeneity.

Figure 3. Funnel Plot



This figure represents a funnel plot used to informally check potential publication bias. In this figure the x-axis represents standardized correlations while the y-axis represents the standard errors. The funnel plot appears relatively symmetric, indicating the absence of potential publication bias.

Begg and Mazumdar (1994) proposed a test for publication bias based on assessing the significance of the correlation between the ranks of effect estimates and the ranks of their variances. The test involves standardizing the effect estimates to stabilize the variances and performing an adjusted rank correlation test based on Kendall's τ. We also applied the formal rank correlation test of Begg and Mazumdar to test statistically for publication bias. Results indicate that sample sizes are not significantly correlated with the standardized effect sizes (r = -0.245, P = 0.243, NS). This suggests that unpublished studies had similar relative effects to those of published studies.

Finally, a "fail-safe" number (Rosenthal, 1991) has been suggested to detect publication bias. This widely used technique estimates the number of studies having zero effect that would have to be published to reduce the overall mean effect size to non-significance. It has been suggested that a fail-safe number should be presented for all meta-analyses, as an aid in the assessment of the degree of confidence that can be placed in the results (Thornton and Lee, 2000). We used both Rosenthal (1991) and Orwin's (1983) formulas to determine the number of "missing" studies required to reduce the current man effect size (0.34) to a small one (0.10). Results indicated that 889 and 29 studies, respectively averaging a zero effect would have to be added to the results of the retrieved findings in order to change the conclusion regarding the magnitude of the effect from large to small (Cooper, 1998). This large fail-safe number indicates that results are robust because it seems unlikely that there would be that many non-significant unpublished studies in the file drawers (Rosenthal, 1991). Thus, the results indicate that new or unpublished studies not included in our meta-analysis do not represent serious threats to the validity of the findings for the bivariate relationship between warning labels and consumers' recall.

More recently, Duval and Tweedie (2000) proposed a trim-and-fill method to estimate the number of "missing" studies due to publication bias. The method is reliant on the symmetric distribution of effect sizes around the "true" effect size if there is no publication bias, and the simple assumption that the most extreme results have not been published. However, simulation studies have found that the trim-and-fill method detects "missing" studies in a substantial proportion of meta-analysis, even in the absence of bias (Sterne et al., 2001).

In conclusion, we can state that despite small differences in characteristics of published and unpublished studies, the unpublished studies reported in our meta-analysis did not differ in terms of effect size from all previous studies. Thus, there was no direct evidence of publication bias. This conclusion, however, is based on only a small number of unpublished studies. Therefore, it is possible that the unpublished results we obtained are not representative of unpublished studies in this research area.

CONCLUSIONS

In this research we conduct a meta-analysis to synthesize and validate the impact of warning labels on consumers' recall and memory by analyzing the empirical results of various studies over a period spanning around 25 years. The results of our meta-analysis indicate that warning labels moderately influence recall. This finding should emphasize to warning label designers that it is imperative to use vivid-enhancing characteristics in designing warning labels if they hope to achieve high levels of recall. Previous research has found that overexposure or "wear-out" is a major problem for any warning label message that consumers are exposed to repeatedly over time (e.g., Strahan et al., 2002). Pictorial warning labels may reduce the wear out effect and enhance recall. For example, it might be easier for a consumer to avoid reading a text message than to avoid seeing a picture of cancerous lungs that covers half of the front of a cigarette pack.

Although our meta-analysis results indicate that consumers moderately remember warning labels, warning label designers may enhance consumers' recall through the design of simple warning messages where the importance of information is made explicit rather than left to the inferences of the consumer. Limiting the number of inferences consumers must make is especially critical because some researchers have found that older consumers have difficulty making appropriate inferences from warning label texts (e.g., Hasher et al., 1987). Previous research indicates also that the presence of difficult words in warning labels may negatively affect recall of warnings by some consumer groups (e.g., Lepkowska-White and Parsons, 2001). Thus, warning labels are supposed to be designed so that a majority of consumers can comprehend them well and use products more safely.

Meta-analyses have several strengths, but thy also suffer from some inherent limitations. First, it must be noted that, though meta-analyses are more generalizable than any one study because the resulting effect size represents a greater variety of primary sample characteristics than can be achieved through a single primary study, by its nature, meta-analysis contains an inherent sampling bias towards quantitative studies that report effect sizes. Thus our research should be considered a summary of the available quantitative research on the effects of warning labels on consumer recall. Second, because of the limited number of studies included in our meta-analysis ($n = 14$), the study may have a limited statistical power. Using a Monte Carlo simulation, Field (2001) found that a meta-analysis should include at least 15 studies, otherwise the Type I error (accepting a false null hypothesis) could be severely inflated. As we were limited by the availability of research studying the relationship between warning labels and consumer recall, future researchers may include more studies that focus on this area. The more studies that are included, the more creditable are the results at representing this investigated domain. Third, some concerns has recently been voiced over the liberal inclusion criteria of some meta-analyses. It has been suggested that when studies suffering major methodological flaws are included in a meta-analysis, the internal validity of the review is jeopardized (Fitzgerald and Rumrill, 2003). Although journal editors want to ensure published studies are those conducted with scientific rigor, future researchers should address this concern as it relates to meta-analytic reviews. Finally, while our meta-analysis shows that consumers can moderately recall information presented in a warning label (r = 0.35, N = 1882), moderating factors have not been analysed in this study. Future research should explore the influence of moderator factors, such as vividness-enhancing characteristics, familiarity with the product, placement of the warning label, the use of color and pictorial features, and the format of warning label information on consumers' recall of its contents.

Appendix: Sample R codes used to conduct the meta-analysis

```
install.packages("metafor")
install.packages("metacor")
install.packages("meta")
library(metafor)
library(metacor)
library(meta)
data.recall <- read.delim("[path location]/recall.txt", header = T, sep = "\t")
data.recall
nbsp.DSL.metacor=metacor.DSL(data.recall$r, data.recall$N, data.recall$Label)
nbsp.DSL.metacor
nbsp.OP.metacor=metacor.OP(data.recall$r, data.recall$N, data.recall$Lbel)
nbsp.OP.metacor
#  Tests of heterogeneity
nbsp.DSLfromMeta=metacor(cor=r, n=N, studlab=Label, data=data.recall)
summary(nbsp.DSLfromMeta)
```

REFERENCES

(References marked with an asterisk indicate studies included in the meta-analysis)

Adams, D., Anthony, C. (1996). Using randomisation techniques to analyse behavioural data. *Animal Behaviour*, 51, 733-738.

Adams, D., Gurevitch, J., & Rosenberg, M. (1997). Resampling tests for meta-analysis of ecological data. *Ecology*, 8, 1277-1283.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.

Andrews, J., Netemeyer, R., & Durvasula, S. (1991). Effects of consumption frequency on believability and attitudes toward alcohol warning labels. *Journal of Consumer Affairs*, 25, 323-338.

Attia, J., Thakkinstian, A., & D'Este, C. (2003). Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *Journal of Clinical Epidemiology*, 56, 297-303.

Bansal-Travers, M., Hammond, D., Smith, P., & Cummings, K. (2011). The impact of cigarette pack design, descriptors, and warning labels on risk perception in the U.S. *American Journal of Preventive Medicine*, 40, 674-682.

*Barlow, T. & Wogalter, M. (1991). Alcohol beverage warning labels in print advertisements. in *Proceedings of the Human Factors Society 35th Annual Meeting*, Santa Monica, CA, 451-455.

Begg, C., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088-1101.

Beltrmini, R. (1988). Perceived believability of warning label information presented in cigarette advertising. *Journal of Advertising*, 17, 26-32.

*Bhalla, G., & Lactovicka, L. (1984). The impact of changing cigarette warning message content and format. in *Proceedings of the Human Factors Society 29th Annual Meeting*, Santa Monica, CA, 305-310.

Borenstein, M. (2005). Software for publication bias. in Rothstein, H., Sutton, A., & Borenstein, M. (Eds.), *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments*. John Wiley & Sons Ltd, New York.

Borland, R., Wilson, N., Fong, G., Hammond, D., Cummings, K.., & Yong, H. (2009). Impact of graphic and text warnings on cigarette packs: Findings from four countries over five years. *Tobacco Control*, 18, 358-364.

Chen, D., & Peace, K. (2013). Applied meta-analysis with R. CRC Press. Boca Raton, FL.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. Academic Press, New York.

Cooper, H. (1998). *Synthesizing Research*, 3rd ed. Sage Publication, Thousand Oaks, CA.

Crespo, A., Cabestrero, R., Grizb, G., & Quiros, P. (2007). Visual attention to health warnings in cigarette advertisements An eye-tracking research between smokers and non-smokers. *Studia Psychologica*, 49, 39-351.

Davis, A. & Rothstein, H. (2006). The effects of the perceived behavioural integrity of managers on employee attitudes: a meta-analysis. *Journal of Business Ethics*, 67, 407-419.

*Desaulniers, D. (1987). Layout, organization, and the effectiveness of consumer product warnings. in *Proceedings of the Human Factors Society 31st Annual Meeting*, Santa Monica, CA, 56-60.

Duval, S., & Tweedie, R. (2000). A non-parametric 'trim and fill' method of assessing publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89-98

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association*, 82, 171-200.

Efron, B. (1979). Bootstrap methods: another look at the jacknife. *Annual Statistician*, 7, 1-26.

Fern, E., & Monroe, K. (1996). Effect size estimates: issues and problems in interpretation. *Journal of Consumer Research*, 23, 89-115.

Field, A. (2001). Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161-180.

Fitzgerald, S., & Rumrill, P. (2003). Meta-analysis as a tool for understanding existing research literature. *Work*, 21, 97-103.

Fleiss, J. (1993). The statistical basis of meta-analysis. *Statistical Methods in Medical Research*, 2, 121-145.

Fox, R., Krugman, D., Fletcher, J., & Fischer, P.(1998). Adolescents' attention to beer and cigarette print ads and associated product warnings. *Journal of Advertising*, 27, 57-68.

*Frantz, J. (1992). *Effect of location, procedural explicitness, and presentation format on user processing of and compliance with product warnings and instructions*. Unpublished Doctoral Dissertation, Department of Industrial and Operations Engineering. University of Michigan.

*Gardner-Bonneau, D., Kabbara, F., Hwang, M., Bean, H., Gantt, M., Hartshorn, K., Howell, J., & Spence, R. (1989). Cigarette warnings: recall of content as a function of gender message content, smoking habits and time. in *Proceedings of the Human Factors Society 33rd Annual Meeting*, Santa Monica, CA, 928-930.

Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Review of Research in Education*, 5, 351-379.

*Goldhaber, G., & deTurck, M. (1988). Effectiveness of warning signs: gender and familiarity effects. *Journal of Products Liability*, 11, 271-284.

Goyal, M., Rajan, S, Essien, E., & Sansgiry, S. (2012). Effectiveness of FDA's new over-the-counter acetaminophen warning label in improving consumer risk perception of liver damage. *Journal of Clinical Pharmacy and Therapeutics*, 37, 681-685.

Gregoire, G., Derderian, F., & LeLorier, J. (1995). Selecting the language of the publications included in meta-analysis: is there a tower of babel bias? *Journal of Clinical Epidemiology*, 48, 195-163.

Grewal, D., Kavanoor, S., Fern, E., Costley, C., & Barnes, J. (1997). Comparative versus non-comparative advertising: a meta-analysis. *Journal of Marketing*, 61, 1-15.

Gurevitch, J., & Hedges, L. (1999). Statistical issues in ecological meta-analysis. *Ecology*, 80, 1142-1149.

Haines-Saah, R., Bell, K., & Dennis, S. (2015). A qualitative contenet analysis of cigarette health warning labels in Australia, Canada, the United Kingdom, and the United States. *American Journal of Public Health*, 105, 61-69.

Hammond, D., McDonald, P., Fong, G., Brown, K., & Cameron, R. (2004). The impact of cigarette warning labels and smoke-free bylaws on smoking cessation: Evidence from former smokers. *Canadian Journal of Public Health*, 95, 201-204.

Hankin, J., Sloan, J., & Sokol, R. (1998). The modest impact of the alcohol beverage warning label on drinking pregnancy among a sample of African-American. *Journal of Public Policy & Marketing*, 17, 61-69.

Harden, A., & Thomas, J. (2005). Methodological issues in combining diverse study types in systematic reviews. *International Journal of Social Research Methodology*, 8, 257-271.

Hasher, R., Zacks, L., Doren, B., Hamm, V., & Attig, M. (1987). Encoding and memory of explicit and implicit information. *Journal of Gerontology*, 42, 418-422.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press, San Diego, CA.

Hsu, L. (2002). Fail-safe Ns for one- versus two-tailed tests lead to different conclusions about publication bias. *Understanding Statistics*, 1, 85-100.

Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting Error and Bias in Research Findings*, 2nd ed., Sage, Thousand Oaks, CA.

Hunter, J., & Schmidt, F. (2000). Fixed effects vs. random effects meta-analysis models: implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8, 275-292.

Ioannidis, J., Trikalinos, T., & Zintzaras, E. (2006). Extreme between-study homogeneity in meta-analyses could offer useful insights. *Journal of Clinical Epidemiology*, 59, 1023-1032.

*Kalsher, M., Clarke, S., & Wogalter, M. (1991). Posted warning placard: effects on college students' knowledge of alcohol facts and hazards. in *Proceedings of the Human Factors Society 35th Annual Meeting*, Santa Monica, CA, 456-460.

*Karnes, E., & Leonard, S. (1986). Consumer product warnings: reception and understanding of warning information by final users. in *Trends in Ergonomics/Human Factors III*, W. Karwowski (Ed.). Elsevier Science Publishers, North Holland, 995-1003.

Kaskutas, L. (1993). Changes in public attitudes toward alcohol control policies since the warning label mandate of 1988. *Journal of Public Policy & Marketing*, 12, 30-37.

Kees, J., Burton, S., Andrews, C., & Kozup, J. (2010). Understanding how graphic pictorial warnings work on cigarette packaging. *Journal of Public Policy & Marketing*, 29, 265-276.

Krugman, D., Fox, R., Fletcher, J., Fischer, P., & Rohas, T. (1994). Do adolescents attend to warnings in cigarette advertising? An eye-tracking approach. *Journal of Advertising Research,* 34, 39-52.

Lacanilao, R., Cash, S., & Adamowicz, W. (2011). Heterogeneous consumer responses to snack food taxes and warning labels. *Journal of Consumer Affairs*, 45, 108-122.

Lehto, M. & Miller, J. (1986). *Warnings: Fundamentals, Design, and Evaluation Methodologies*. Fuller Technical Publications, Ann Arbor, MI.

Lepkowska-White, E., & Parsons, A. (2001). Comprehension of warning and resulting attitudes. *Journal of Consumer Affairs*, 35, 278-294.

Light, R., & Pillemer, D. (1984). *Summing Up: The Science of Reviewing Research*. Harvard University Press, Cambridge, MA.

Loken, B., & Howard-Pitney, B. (1988). Effectiveness of cigarette advertisements on women: An experimental study. *Journal of Applied Psychology*, 73, 378-382.

MacKinnon, D., Nohre, L., Pentz, M., & Stacey, A. (2000). The alcohol warning and adolescents: 5-year effects. *American Journal of Public Health*, 90, 1589-1594.

MacKinnon, D., Nohre, L., Cheong, J., Stacey, A., & Pentz, M. (2000). Longitudinal relationship between the alcohol warning and alcohol consumption. *Journal of Studies on Alcohol*, 62, 221-226.

Manly, B. (1991). *Randomization and Monte Carlo Methods in Biology*. Chapman & Hall. New York.

Marin, G. (1997). Changes across 3 years in self-reported awareness of product warning messages in a Hispanic community. *Health Education Research*, 12, 103-116.

Mazis, M., Morris, L., & Swasy, J. (1996). Longitudinal study of awareness, recall, and acceptance of alcohol warning labels. *Applied Behavioral Science Review*, 4, 111-120.

Miller, C., Hill, D., Quester, P., & Hiller, J. (2009). Impact on the Australian Quitline of new graphic cigarette pack warnings including the Quitline number. Tobacco Control, 18, 235-237.

Moller, A., & Thornhill, R. (1998). Bilateral symmetry and sexual selection: a meta-analysis. *American Naturalist*, 151, 174-192.

Moodie, C., MacKintosh, A., & Hammond, D. (2009). Adolescents' response to text-only tobacco health warnings: Results from the 2008 UK Youth Tobacco Policy Survey. *European Journal of Public Health*, 20, 463-469.

Morris, N., Gilpiin, D., Lenos, M., & Hobbs, R. (2011). Interpretations of cigarette advertising warning labels by Philadelphia Puerto Ricans, *Journal of Health Communications*, 16, 908-922.

Muggli, M., Pollay, R., Lew, R., & Joseph, A. (2002). Targeting of Asian Americans and Pacific Islanders by tobacco industry: Results from the Minnesota Tobacco Document Depository. *Tobacco Control*, 11, 201-209.

*Nohre, L., MacKinnon, D., Stacey, A., & Pentz, M. (1999). The association between adolescents' receiver characteristics and exposure to the alcohol warning label. *Psychology & Marketing*, 16, 245-259.

Orwin, R. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistic*s, 8, 157-159.

Orwin, R., Schucker, R., & Stokes, R. (1984). Evaluating the life cycle of a product warning: Saccharin and diet soft drinks. *Evaluation Review*, 801-822.

*Otsubo, S. (1988). *Effectiveness of warning signs on consumer products*. Unpublished Masters Thesis. Department of Psychology. California State University at Northridge.

Palmatier, R., Dant, R., Grewal, D., & Evans, K. (2006). Factors influencing the effectiveness of relationship marketing: a meta-analysis. *Journal of Marketing*, 70, 136-153.

Petera, E., Romer, D., Slovic, P., Jamieson, K., Wharfield, L., Mertz, C., & Carpenter, S. (2007). The impact and acceptability of Canadian-style cigarette warning labels among U.S. smokers and non-smokers. *Nicotine & Tobacco Research*, 9, 473-481.

Petitti, D. (1994). *Meta-analysis, Decision Analysis, and Cost-Effectiveness Analysis*. Oxford University Press, New York.

Rosenberg, M., Adams, D., & Gurevitch, J. (2000). *MetaWin: Statistical Software for Meta-analysis, Version 2.0.* Sinauer, Sunderland, MA.

Rosenthal, R. (1991). *Meta-analytic procedures for Social Research*. Russell Sage Foundation, New York.

Rosenthal, R., & DiMatteo, M. (2001). Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-82.

Ruiter, R., & Kok, G. (2005). Saying is not (always) doing: Cigarette warning labels are useless. *European Journal of Public Health*, 15, 329-336.

Scholes-Balog, K., Heerde, J., & Hemphill, S. (2012). Alcohol warning labels: Unlikely to affect alcohol-related beliefs and behaviors in adolescents. *Australia and New Zealand Journal of Public Health*, 6, 524-529.

Sterling, T., Rosenbaum, W., & Weinkam, J. (1995). Publication decisions revisited: the effect of outcome of statistical tests on the decision to publish and vice versa. *American Statistician*, 4, 108-112.

Sterne, J., Egger, M., & Davey, G. (2001). Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal*, 323 (July, 14), 101-105.

Sterne, J., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, 53, 1119-1129.

Stewart, D., & Martin, I. (1994). Intended and unintended consequences of warning messages: a review and synthesis of empirical research. *Journal of Public Policy & Marketing*, 13, 1-19.

Stockley, C. (2001). The effectiveness of strategies such as health warning labels to reduce alcohol-related harms: an Australian perspective. *International Journal of Drug Policy*, 12, 153-166.

Strahan, E., White, K., Fong, G., Fabrigar, L., Zanna, M., & Cameron, R. (2002). Enhancing the effectiveness of tobacco package warning labels: a social psychological perspective. *Tobacco Control*, 11, 183-190.

*Strawbridge, J. (1986). *The influence of position, highlighting, and embedding on warning effectiveness*. Unpublished Masters Thesis. Department of Industrial and Systems Engineering, California State University at Northridge.

Takkouche, B., Cadarso-Suarez, C. & Spiegelman, D. (1999). Evolution of old and new tests of heterogeneity in epidemiologic meta-analysis. *American Journal of Epidemiology*, 150, 206-21.

Tam, T., & Greenfield, T. (2010). Do alcohol warning labels influence men's and women's attempts to deter others from driving when intoxicated? *Human Factors and Ergonomics in Manufacturing & Service Industries*, 20, 538-546.

Tang, J., & Liu J. (2000). Misleading funnel plot for detection of bias in meta-analysis. *Journal of Clinical Epidemiology*, 53, 477-484.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: its causes and consequences. *Journal of Clinical Epidemiology*, 53, 207-218.

Thrasher, J., Pérez-Hernández, R., Arillo-Santillán, E., & Barrientos-Gutiérrez, I. (2012). Impacto de las advertencias con pictogramas en las cajetillas de cigarrillos en México: Resultados de una encuesta en fumadores de Guadalajara. *Salud Pública de México*, 54, 254-263.

White, V., Webster, B., & Wakefield, M. (2008). Do graphic health warning labels have an impact on adolescents' smoking-related beliefs and behaviors? *Addiction*, 103, 1562-1571.

Wilkinson, C., &Room, R. (2009). Warnings on alcohol containers and advertisements: International experience and evidence on effects. *Drug and Alcohol Review*, 28, 426-443.

*Wogalter, M., Kalsher, M., & Racicot, B.(1992). The influence of location and pictorials on behavioural compliance to warnings. in *Proceedings of the Human Factors Society 36th Annual Meeting*, Santa Monica, CA, 1029-1033.

Yong, H., Borland, R., Thompson, M., Fong, G., Thrasher, J., Nagelhout, G. & Hammond, D. (2014). Mediational pathways of the impact of cigarette warning labels on quit attempts. *Health Psychology*, 33, 1410-1420.

*Young, S., & Wogalter, M. (1990). Comprehension and memory of instruction manual warnings: conspicuous print and pictorial icons. *Human Factors*, 32, 637-649.

BIOGRAPHY

Mohamed M. Mostafa has received a PhD in Business and Marketing from the University of Manchester, UK and a PhD in International Affairs from the University of St. Gallen, Switzerland. He has received several masters' degrees in fields as diverse as Functional Brain Imaging (Brunel, UK), Applied Statistics (Northern Colorado, USA), Social Science Data Analysis (Essex, UK), Translation Studies (Portsmouth, UK), French Culture and Civilization (Middlebury College, USA) and Marketing (Port Said, Egypt). He is currently a full professor of Marketing and Decision Sciences at the Gulf University for Science and Technology, Kuwait, having previously been employed as a visiting professor of marketing at Auburn University, U.SA, and at other universities in Egypt, Cyprus, Jordan, United Arab Emirates (UAE), France, Bahrain, Turkey, and Portugal. His current research interests include artificial Intelligence applications in marketing/business, neuro-marketing, social marketing and business efficiency. He has published over 70 papers in several leading academic peer reviewed journals, including *Psychology & Marketing*, *Journal of Managerial Psychology*, *Social Indicators Research*, *Journal of Promotion Management*, *International Journal of Consumer Studies*, *International Journal of Information Management, International Journal of Retail and Distribution Management, International Journal of Health Care Quality Assurance*, *Health Marketing Quarterly, International Journal of Business Performance Management, Expert Systems with Applications, Computational Statistics and Data Analysis, International Journal of Intelligent Computing & Cybernetics, Applied Mathematics & Computation* and the *Journal of Economic Studies*. He has als presented numerous papers at professional conferences worldwide. Email: mostafa@usa.com.