

Pictorial Health Warnings:
A Review of Research Evidence

Research Report Prepared for

Ministry of Health

By

Professor Janet Hoek

Professor Philip Gendall

April, 2005

Table of Contents

Executive Summary	i
1. Introduction	1
2. Criticisms of the MoH Consultation Document	3
2.1 Tobacco Industry Submissions and Reports	3
2.1.1 BAT Submission	3
2.1.2 Scientific Critique by Dr M Dixon	3
2.1.2 IT Submission and Supporting Documents	3
2.1.3 Supporting Statement by Mr Andrew Brown QC	5
2.1.4 Criticisms made by Associate Professor Sarah Todd	6
2.2 Analysis of Arguments Presented in BAT and IT Submissions	7
2.2.1 Existing High Levels of Awareness	7
2.2.2 Lack of Evidence of Behavioural Change	7
2.2.3 Value of Longitudinal Studies	8
2.2.4 Alternative Research Approaches	9
2.2.5 Methodological Criticisms	10
2.2.6 Evaluation of Criticisms Levelled at Specific Research Papers	12
3. Conclusions	46
3.1 Criticisms of the MoH Consultation Document	46
3.2 Summary Comments	50
Bibliography	50

Executive Summary

In July 2004, the MoH published a consultation document that summarised and discussed research relating to the effectiveness of pictorial health warnings (Ministry of Health, 2004). British American Tobacco's (BAT) and Imperial Tobacco's (IT) submissions on the consultation document raised questions about the robustness of research reported on in the Ministry of Health (MoH) consultation document. In particular, the tobacco industry submissions argued that the relationship between pictorial health warnings and smoking-related behaviours remained unproven. More specifically, the BAT and IT documents questioned whether pictorial health warnings would promote cessation among smokers, or inhibit trial and the development of a smoking habit among current non-smokers. Both companies argued strongly that evidence of likely behaviour change was a necessary condition that must be met before the case for requiring pictorial health warnings could be established.

To consider the issues raised by BAT and IT in more detail, the MoH commissioned an independent review of the research evidence included in their consultation document. The review's purpose was to analyse the MoH's interpretation and use of the research evidence relied on in the consultation document, and to identify any limitations of this research. The review addressed the following issues:

- Alleged flaws in the methodology of studies cited in the discussion document;
- The alleged lack of quality studies cited;
- The researchers' allegedly inappropriate reliance on attitudinal changes as an indicator of the effectiveness of pictorial health warnings;
- The argument that the only true measure of effectiveness is direct measurement of behavioural change that can be attributed to exposure to health warnings.
- The practicality of designing and conducting the type of research advocated as necessary by BAT and IT, and
- The appropriate form and extent of research required for intervention design and evaluation, and appropriate measures of effectiveness for assessing the likely outcome, should pictorial health warnings be introduced.

We have found that:

- The flaws alleged to have affected the studies reported in the MoH consultation document are typically often presented in general terms without detailed evidence of their effects. As far as we can tell, the surveys discussed in the MoH consultation document were well designed and conducted. We have seen no evidence that the estimates reported have been materially affected by the errors outlined in the tobacco industry submissions, although it is not possible to assess all the alleged errors.

- Much of the research cited used changes in awareness or other cognitive variables as outcome measures. These measures cannot be assumed to lead to behaviour, nor do they necessarily offer insights into respondents' future behaviour.
- Arguments that evidence of behaviour change is necessary to support proposals to introduce pictorial warning labels are logically flawed as they pre-suppose the introduction of the measures they propose to assess.
- Although behaviour change is a desired outcome of proposals to introduce pictorial warning labels, we believe it would be extremely difficult to undertake research that could isolate the specific contribution of pictorial warning labels to smoking cessation or non-initiation, and that would be accepted by all parties.
- In the absence of behavioural measures, it is nevertheless possible to compare alternative warning labels and to obtain estimates of their likely effects on respondents' behaviour.
- If pictorial warning labels were introduced, appropriate evaluation measures would depend on the objectives they were intended to realise. Communication objectives could be assessed using a similar methodology to that employed by Health Canada. Behavioural objectives could be assessed using a phased study that incorporated control sub-samples.

1. Introduction

As part of its on-going strategy to reduce the consumption of tobacco products, and thus the adverse health consequences that result from smoking, the New Zealand government is considering whether to mandate the use of pictorial health warnings on tobacco products. In July 2004, the Ministry of Health (MoH) published a consultation document that summarised and discussed research relating to the effectiveness of pictorial health warnings (Ministry of Health, 2004). The consultation document included four options:

Option 1: Non-regulatory methods (mass media, education in schools, etc)

Option 2: The use of inserts inside tobacco packs

Option 3: Refreshing the current warnings

Option 4: Adopting pictorial warnings in addition to text

In addition, the consultation document outlined a series of options relating to the location and size of health warnings, the number and rotation of health warnings, the use of te reo Māori health warnings, and the attribution of health warnings.

British American Tobacco (BAT) and Imperial Tobacco (IT) made submissions on the consultation document; among other concerns, both companies raised questions about the robustness of research reported on in the MoH consultation document. Their submissions alleged several deficiencies affected the research used by the MoH. In particular, the tobacco industry submissions argued that the relationship between pictorial health warnings and smoking-related behaviours remained unproven. More specifically, the BAT and IT documents questioned whether pictorial health warnings would promote cessation among smokers, or inhibit trial and the development of a smoking habit among current non-smokers. Both companies argued strongly that evidence of likely behaviour change was a necessary condition that must be met before the case for requiring pictorial health warnings could be established.

To consider the issues raised by BAT and IT in more detail, the MoH commissioned an independent review of the research evidence included in their consultation document. The review's purpose was to analyse the MoH's interpretation and use of the research evidence relied on in the consultation document, and to identify any limitations of this research. More specifically, we were required to report on the following issues:

- Alleged flaws in the methodology of studies cited in the discussion document;
- The alleged lack of quality studies cited;
- The researchers' allegedly inappropriate reliance on attitudinal changes as an indicator of the effectiveness of pictorial health warnings;

- The argument that the only true measure of effectiveness is direct measurement of behavioural change that can be attributed to exposure to health warnings.

In addition, we were asked to consider:

- The practicality of designing and conducting the type of research advocated as necessary by BAT and IT, and
- The appropriate form and extent of research required for intervention design and evaluation, and appropriate measures of effectiveness for assessing the likely outcome, should pictorial health warnings be introduced.

We begin by summarising the criticisms levelled at the research reported on in the MoH document; we then evaluate the logic of those criticisms and present an analysis of the extent to which we believe they are justified.

2. Criticisms of the MoH Consultation Document

2.1 Tobacco Industry Submissions and Reports

2.1.1 BAT Submission

The BAT submission noted that awareness of the health risks of smoking is high and argued that new warnings are unlikely to increase this level of awareness, and are therefore not justifiable. In addition, BAT argued there is no evidence that pictorial warnings change smokers' behaviour or reduce the likelihood that non-smokers will develop a smoking habit. BAT claimed that the Canadian research referred to in the consultation document focussed on attitudinal rather than behavioural change, and stated that the former is not necessarily predictive of the latter. Because BAT argued that the benefits of any proposed changes should be demonstrated before those changes are introduced, they alleged the lack of evidence relating to behaviour meant there is insufficient justification for the introduction of pictorial health warnings.

BAT also alleged that pictorial health warnings may lead to adverse effects, such as reactance, and that they may shame smokers, an outcome BAT believes is inappropriate. They further alleged that pictorial health warnings would wear out quickly, and thus have limited effectiveness.

BAT's submission also questioned the process followed by the MoH in preparing the consultation document. In paragraphs 113 onwards, BAT noted the need for the MoH to follow the principles of natural justice and to undertake meaningful consultation. BAT believes such consultation should involve consideration of the full range of evidence; acknowledgement of any methodological limitations affecting the conclusions drawn in research relied upon by the MoH, and demonstration of a lack of bias.

Overall, BAT argued that New Zealanders are already well aware of the health risks caused by smoking and that a change in the format of on-pack warnings is unlikely to increase this. BAT also criticised the MoH's alleged reliance on attitudinal studies, which BAT believes do not necessarily enable prediction of behavioural outcomes. As a result, BAT concluded that there is little or no evidence that the introduction of pictorial health warnings will bring any benefits.

As well as documenting these general concerns, the BAT submission outlined a series of more specific concerns about research papers drawn on in the MoH consultation document. We evaluate these criticisms in section 2.2 below.

2.1.2 Scientific Critique by Dr M Dixon

BAT's submission included a critique of the MoH consultation document undertaken by Dr M Dixon, a senior scientific advisor to BAT. Dixon's report reviewed all aspects of the consultation document;

section 3 of his report dealt specifically with health warnings. Our review draws only on section 3 of Dixon's report, which reviewed papers cited in the Allen & Clarke (A&C) report and consultation document. Dixon made several criticisms of these papers and the manner in which they have been used. These criticisms included:

- The MoH's heavy reliance on attitudinal rather than behavioural research. In making this criticism, Dixon quoted Dr John Liefeld who differentiated between Type 1 (behavioural) and Type 2 (cognitive and affective) research.
- The MoH's alleged mis-quotation of the Environics research, which he argued provides strong evidence that the introduction of pictorial health warnings has not increased the proportion of smokers considering quitting-related behaviours. Furthermore, he argued that much of the Canadian research did not involve baseline surveys that would enable an estimation of whether any differences observed could be attributed to pictorial health warnings. As a result, he concluded that attempts to draw causal inferences about the effect of pictorial health warnings on behaviour are inherently flawed. In addition, he stated that follow-up surveys by Health Canada that could enable conclusions to be drawn about the effect pictorial health warnings have had on smokers' behaviour had not been publically released.
- The MoH's lack of emphasis on Gospodinov and Irving's work, which he argued found pictorial health warnings had not significantly decreased smoking prevalence or cigarette consumption.
- The MoH's assumption that changes in smokers' attitudes, intentions and behaviour are attributable to the introduction of pictorial health warnings when a variety of factors, of which pictorial health warnings are only one, may have contributed to the outcomes observed.
- The MoH's pre-determined beliefs about the direction of causality, which could be the opposite of that assumed.
- The lack of evidence provided by the MoH to support their conclusion that current text-based health warnings have suffered from wearout and are less effective as a consequence.

2.1.3 IT Submission and Supporting Documents

Imperial Tobacco's submission made many of the same points outlined in BAT's submission. In particular, IT argued that the MoH had not established that the introduction of pictorial health warnings would alter intended or actual behaviour with respect to smoking. IT also questioned whether a direct and unambiguous link between pictorial health warnings, attitudes towards smoking, and smoking behaviours exists, and whether recall is a valid predictor of future behaviour.

IT noted concern that pictorial health warnings may vilify and shock smokers, and have potentially adverse effects on children and teenagers. Given this alleged outcome, IT questioned whether health warnings should go beyond the provision of information. IT outlined a range of adverse consequences it suggested may follow the introduction of pictorial warnings; these include the use of sleeves to cover the warning images and the growth of a black market for packs that do not feature pictorial health warnings. IT also suggested that pictorial health warnings may promote smoking

among those youth prone to engaging in acts of rebellion (although Barwick *et al* 1995 found little evidence to support this alleged effect). IT challenged the need for further regulatory interventions, since statistics the company uses show that tobacco consumption is already declining more rapidly than in countries where pictorial health warnings have been introduced.

2.1.4 Supporting Statement by Mr Andrew Brown QC

Part of IT's submission included a statement by Mr Andrew Brown QC. Brown canvassed a range of legal issues in his statement, and made some specific references to pictorial health warnings. In particular, Brown argued that the consultation document lacked evidence that pictorial health warnings would result in a material behavioural outcome that could be directly linked to the change in pictorial health warnings. In paragraphs 18-21 of his statement, Mr Brown questioned the relationship between recall, attitudes and behaviour. He elaborated on these comments in paragraphs 19 onwards of his submission, where he referred to a review prepared on behalf of IT by Associate Professor Sarah Todd. In summarising key aspects of Todd's review, he noted she found that:

- Few papers attempted to examine associations between respondents' attitudes and behaviour and their exposure to pictorial health warnings; as a result, it is alleged that the material effects of health warnings have not been clearly demonstrated.
- Few of the papers she reviewed reported a link between recall (or awareness of the health warnings and their content) and smoking cessation or inhibition. As a result, it is alleged that strategies designed to increase awareness of the risks associated with tobacco consumption are unlikely to affect smokers' behaviour or decrease the probability that vulnerable non-smokers will commence smoking. Furthermore, several studies note that awareness of the adverse health consequences of tobacco consumption is already high, thus the need for additional measures designed to reinforce this awareness has allegedly not been established.
- More specific criticisms raised by Todd include an alleged lack of external validity, failure to consider the addictive nature of smoking; a self-reporting bias on respondents' part; social desirability bias; non-response bias; interviewer bias; reliance on studies with small samples; use of unrepresentative samples; and "testing" effects.

In summary, Brown argued that the need for additional warnings has not been demonstrated, that the relationship between pictorial health warnings and smoking cessation, reduction, or inhibition has not been established and that, even if a need for further information did exist, more appropriate communication channels should be used. Furthermore, he argued that the evidence relied on in the consultation document is compromised by the errors outlined above, and that any conclusions advanced on the basis of this evidence are inevitably flawed.

2.1.5 Criticisms made by Associate Professor Sarah Todd

Associate Professor Todd was commissioned by IT to critique the research evidence relied on by the MoH in its consultation document. Her statement elaborated on criticisms documented in IT's submission, raised by BAT, and referred to by Brown. Todd raised general points about the research evidence presented in the MoH consultation document before critiquing several of the research papers individually. Her general criticisms included the following points:

- There are few clear links between pictorial health warnings and behaviour or attitude change, or between recall, agreement with health statements about smoking, or increases in legibility of health warnings and behaviours such as cessation or non-initiation.
- The specific effects of health warnings are difficult to isolate as these are normally part of a wider campaign to reduce the prevalence of smoking or tobacco consumption.
- The definition of "effectiveness" used in the research relied upon by the MoH varied and, at times, was ambiguous.

Todd questioned various methodological aspects of the research and alleged that the sampling frames used in some studies produced samples with insufficient numbers of respondents aged 18 years and under. She also questioned the accuracy and reliability of self-reported data and suggested that some constructs were inadequately estimated. In addition, she alleged that a variety of errors, including social desirability error and measurement error, could have inflated estimates of quitting, where this was measured. She also argued that low response rates could have introduced non-response error into the estimates reported. In addition, she alleged the external validity of some studies was low and questioned the generalisability of the conclusions. Todd noted a dearth of longitudinal studies, which she argued were necessary to establish the longer term behavioural consequences of regulatory interventions. As a more general point, she argued that research relied upon by the MoH was undertaken by researchers whom she did not regard as disinterested because of their involvement in health-related areas, or their receipt of funding by groups regarded as opposed to the tobacco industry. After making these broad points, Todd critically evaluated 13 papers referred to in the MoH consultation document.

2.2 Analysis of Arguments Presented in BAT and IT Submissions

2.2.1 Existing High Levels of Awareness

Both the IT and BAT submissions argued that, as public awareness of the health risks associated with smoking is high, there is no demonstrable need for interventions designed to boost this awareness. However, awareness is not the only relevant variable to be considered in this assessment. Barwick *et al* (1995) noted the difference between awareness of a risk, belief in the reality of that risk, and the extent to which people view that risk as personally relevant. This implies that even if the argument that pictorial health warnings are unlikely to increase aggregate awareness of the health risks of smoking is correct, they may still be justifiable if they increase the believability of those risks. Similarly, if pictorial health warnings make the risks of smoking more salient and relevant, their introduction may also be justifiable, according to Barwick *et al*'s reasoning.

2.2.2 Lack of Evidence of Behavioural Change

The IT and BAT submissions argued strongly that the MoH must justify any proposed introduction of pictorial health warnings by demonstrating these would reduce consumption, promote quitting, and inhibit smoking initiation. However, paragraph 52 of the BAT submission noted that the role of health warnings is to reinforce broader education campaigns. This argument recognises that health warnings typically constitute part of a broader intervention programme (a point made also in IT's submission).

The consultation document acknowledged the difficulty of establishing a causal relationship between regulatory interventions and behavioural outcomes. As Dixon, Brown and Todd noted, regulatory interventions typically occur as part of a wider range of measures designed to shape and reinforce particular behaviours. Even where interventions are introduced as one-off measures, they are often accompanied by media commentary, which may contribute to any observed effects.

This suggests that obtaining evidence of a direct behavioural link between health warnings and subsequent behaviour may be impossible. The difficulty of controlling all other aspects of a broader smoking cessation or inhibition programme means research attempting to isolate the contribution of one component of that programme would almost inevitably be vulnerable to allegations that other variables had not been adequately controlled, and that confounding factors had influenced the analysis.

Application of the principle that interventions are not justified until they are established to have a behavioural outcome (demonstrated through longitudinal research) creates a logically impossible situation, since it pre-supposes the introduction of the very measures it argues should be evaluated. It is not possible to assess the effects of pictorial health warnings on behaviours such as penetration and consumption until the warnings have been in place for some time. Although, in this particular

case, it would be possible to refer to Canadian research, the general argument that the behavioural benefits of interventions must be established before the intervention can be introduced is logically unsustainable.

2.2.3 Value of Longitudinal Studies

Longitudinal research can be undertaken in a variety of ways and the submissions are not specific about the type of longitudinal research the tobacco industry believes is desirable. They may take the form of repeated cross-sectional studies (the data used by Gospodinov and Irvine, 2003 and 2004 take this form), although these data offer fewer insights into the causal role played by changes in the environment. Alternatively, longitudinal research may use a prospective design, where a panel of individuals is questioned at pre-designated time intervals. Because the same individuals are interviewed over a time period, it is possible to detect changes in their behaviour and to link these to environmental changes (Borland & Hill, 1997; Hammond *et al* 2003; 2004 used this approach).

However, while longitudinal research can be valuable in identifying longer term trends in behaviour, it is not a panacea to the problems that allegedly affect cross-sectional studies. Longitudinal studies require a panel committed to an on-going involvement in the research project. These panels suffer from attrition over time and the recruitment surveys used to establish the panel often achieve low response rates. Over time, the problem of non-response that Todd argued affected studies relied upon in the consultation document may also affect longitudinal studies.

Like cross-sectional studies, longitudinal research projects may also be affected by extraneous variables that complicate researchers' ability to isolate the contribution of a specific element of an intervention. Because data are collected at specific time points, the course of events between these time points is not known with certainty. Unless the external environment was carefully controlled, longitudinal studies may be no better able than cross-sectional studies in quantifying the contribution made by pictorial health warnings, or any other aspect of a regulatory initiative.

Members of a panel are interviewed on a number of occasions over the research period. If the questions used in successive interview waves remain the same, it is possible that responses given in one wave will influence those provided subsequently. Similarly, respondents' very participation in the research may also condition their responses over time.

Despite their support for longitudinal research, both Dixon and Todd criticised studies that employed this approach. Todd alleged "testing" effects had biased estimates obtained by Borland & Hill (1997) and both she and Dixon criticised Hammond *et al* for attributing to warning labels changes they argued could be explained by other variables. These criticisms highlight the difficulty of undertaking longitudinal research and suggest findings from this will not be immune to criticism.

Econometric studies that have investigated the relationship between aggregate consumption and regulatory interventions have been characterised by methodological disputes where researchers fail to agree on the specification of the model used to test the relationship. The exchange between Laugesen & Meads (1991) and Stewart (1992), and the criticisms of Chetwynd *et al's* (1988) research by Boddewyn (1989), illustrate the lack of consensus that affects this work.

Overall, although longitudinal research may enable an assessment of causal relationships, any reported relationships are likely to be the topic of considerable debate because of the difficulty in controlling all other sources of influence. As a result, while we believe longitudinal data could provide useful information, we do not believe the decision of whether to introduce pictorial health warnings can only be informed by such data.

2.2.4 Alternative Research Approaches

In general, we believe research designed to predict behavioural responses is preferable to research that merely obtains measures of affect or cognition, particularly where the objective of an intervention is to shape specific behaviours. However, in this case, establishment of causality would, as the IT and BAT submissions suggest, require longitudinal studies. Yet, these longitudinal studies and the need to measure behavioural outcomes would require the introduction of the package designs that IT and BAT oppose. That is, acceptance of the tobacco companies' submissions about the need to demonstrate a behavioural relationship between pictorial health warnings and smoking behaviours would mean that the evidence required could never be collected.

There are two possible solutions to these problems:

1. Longitudinal studies with behaviour outcomes in countries that have introduced pictorial health warnings. However, the generalisability of these outcomes to New Zealand may be questionable and the problems affecting longitudinal research generally may also lead to disputes over how the findings should be interpreted.
2. Well designed cross-sectional studies conducted in New Zealand that compare the likely effects of alternative warnings on a range of behavioural intention, affective and cognitive variables.

Rather than attempt to measure whether pictorial warnings prompt smokers to cease, or inhibit current non-smokers from developing a habit, research that evaluated alternative warning labels would be an alternative decision-making tool. A comparison of the current warnings and the proposed pictorial warnings could explore attitudinal variables, including the credibility of the warning, its salience and personal relevance. It could also use established methodologies for predicting a range of near-behavioural outcomes, such as the likelihood of phoning the Quitline, reducing consumption, and quitting. Because this type of research could use a variety of "effectiveness" measures, it would

enable a detailed comparison of proposed and existing interventions. If the pictorial health warnings elicited significantly higher scores on key variables, their introduction could be justified on the basis that they were more likely to be effective than the current warnings.

2.2.5 Methodological Criticisms

Todd identified several methodological issues she believes have been inadequately considered and suggested these could affect the emphasis placed on particular research studies. We discuss the concerns she has noted below (where these have not already been discussed in earlier sections).

Ambiguous and Varied Measures of Effectiveness

Our initial assessment of the research papers referred to in the consultation document suggests that "effectiveness" was not always clearly defined and that the outcome variables examined have differed from one study to another. However, because a range of variables can influence the impact of health warnings, and thus their ability to influence behaviour, the use of different outcome variables is not necessarily problematic. Nevertheless, the range of variables examined in the research literature indicates that caution is required when comparing the results of different studies, and when drawing conclusions about what the research findings establish.

Social Desirability Error

Social desirability error occurs when respondents do not provide honest answers when asked about their attitudes, beliefs or behaviour, but instead offer responses they believe will show them in a better light. Where behaviours that have a social stigma attached to them, such as smoking, drink-driving, or child abuse, are investigated there is always a risk that respondents will not admit to performing these behaviours or, if they do admit to them, that they will over-estimate the likelihood that their behaviour will change in the future.

If research does not include an existing measure as a control, it may be difficult to estimate the effect of social desirability error. However, comparative research that evaluates proposed new measures against an existing measure is less affected by social desirability error because the research examines the relative effectiveness of the measures rather than estimating their absolute effect on a population's behaviour.

It is important to note that any data involving self-reports, whether it is part of a cross-sectional project or a longitudinal study, may be affected by social desirability error. We have paid particular attention to research Todd alleged has been affected by social desirability error.

Non-Response Error

Non-response error occurs when individuals who decline to take part in a survey, or who are not available or able to participate, differ from those who did respond. For this reason, researchers attempt to maximise survey response rates, since the higher the response rate the lower the potential for non-response error. However, surveys with high response rates do not always avoid non-response error, just as surveys with low response rates are not always affected by non-response error. Because survey objectives often involve estimating parameters for which there is no known population value, it is normally not possible to gauge the presence of non-response error, or the extent to which this may have affected the survey estimates.

It is sometimes possible to use other information in a sampling frame to compare a sample's characteristics to those of the population from which the sample was drawn. Researchers attempting to estimate the attitudes of the general public may compare the demographic characteristics of their sample to Census data. However, the fact that respondents' demographic profile may match that of a wider population does not mean that estimates of the sample's attitudes, knowledge or behaviour will be accurate projections of the population values for these variables. For this reason, non-response error remains very difficult to quantify.

External and Internal Validity

External validity occurs when researchers can generalise the findings of their research to a wider population. Research designs that draw on behavioural settings, or that use stimuli likely to be found in these settings, typically have high external validity. However, these designs may not be able to control all the variables found in "real" settings, thus they may lack internal validity. By contrast, experimental research conducted in a less natural setting is better able to control extraneous variables, but may lack external validity if the research situation does not parallel important aspects of the environment within which the behaviour of interest question normally occurs.

Research Independence

Todd noted that health researchers or "anti-tobacco groups" have undertaken much of the research into the effectiveness of regulatory consumptions on tobacco consumption and related behaviours. She argued that this area would benefit from more independent research. An unfortunate implication of these comments is that health researchers, or those with affiliations to tobacco control groups, cannot be relied upon to undertake rigorous research, or to report their research findings fairly and appropriately. Todd presents no evidence to support the contention that any of the research reported upon has been deliberately biased in its design or interpretation. To test the validity of her concerns, we have paid particular attention to the researchers' backgrounds (where information on this is available), the design of the research they have undertaken, and the extent to which the conclusions advanced are supported by the research findings.

2.2.6 Evaluation of Criticisms Levelled at Specific Research Papers

In this section we examine a range of papers drawn on in the Allen & Clarke report, and relied upon in the MoH consultation document or which have since become available. The work is grouped geographically, to enable some consideration of the research in different regulatory environments. This structure follows the general structure of the consultation document and the Allen & Clarke report that informed this.

Canadian Research

Gospodinov and Irvine 2003 and 2004

We discuss these two papers together, since the 2003 paper is akin to a working paper that was developed further before being published in 2004. Gospodinov and Irvine examined two specific behavioural variables: the prevalence of smoking (i.e., the proportion of the population that meets the definition of a smoker) and consumption (i.e., the amount of tobacco consumed). The researchers criticised research undertaken by Environics, which they argued failed to link the measures tested with the specific outcomes of interest. They also raised more general concerns about the lack of longitudinal studies, the low response rates achieved in some studies, and the variety of methods employed.

To address their concerns over the lack of behavioural research, Gospodinov and Irvine drew upon two waves of a Statistics Canada survey. The first wave included data from the 6 months prior to the January 2001 introduction of pictorial health warnings while the second wave included data from the 6 months that followed the change in regulation. The dataset included over 20,000 cases, of which around 5000 met the definition of a smoker.

Gospodinov and Irvine used a series of advanced multi-variate (regression) models to analyse the data. In their 2003 paper, they reported that although demographic and price variables were significant in both models they developed, the variable used to assess the impact of changes in the health warnings was not significant. While they found evidence that tobacco consumption had dropped by around 4 cigarettes per week from the first period to the second, they suggested this was likely to be part of a general trend towards declining consumption.

The analysis presented in the 2004 paper differed slightly from that presented in the 2003 paper. Gospodinov and Irvine reported that average prevalence of smoking declined from 27.6% to 26.8%; subsequent re-analysis lead them to re-iterate their earlier conclusion that the change in warning label had not affected prevalence. However, while the change observed may not have been statistically significant, it is nevertheless in the direction expected if pictorial warning labels had

affected behaviour. Weekly consumption declined from 25.3 to 22.6 cigarettes, a decline of just over two cigarettes per week. The analysis presented in the 2004 paper found that this decline was large and significant, although Gospodinov and Irvine argued the result was surprising, given the prevalence finding. They concluded that some caution is required in interpreting the latter findings, which they suggested did not necessarily imply the success of the new policy initiative.

Both the prevalence and consumption results indicate a decline in the proportion of smokers and the number of cigarettes they consume. Although only the consumption finding reached accepted levels of significance, we suggest that the difference between the prevalence and consumption results may be less surprising than Gospodinov and Irvine have noted. One effect of the new health warnings may be to promote a reduction in the number of cigarettes consumed; this would explain why the prevalence result remained similar, but the consumption declined. Gospodinov and Irvine did not explain why they found it surprising that health warnings would promote a reduction in consumption rather than cessation. Their argument overlooks the highly addictive nature of smoking and does not recognise that some smokers may wean themselves off their dependency rather than quit immediately.

We note also that changes in regulatory interventions may take some time to affect prevalence rates. The data used to test the effectiveness of the pictorial health warnings were collected immediately after the changes had been introduced and it is arguable that changes in prevalence should be measured over a longer time period. Intermediary behaviours, such as calls to a Quitline (or equivalent), or requests for and sales of nicotine replacement therapies, may have provided more detailed insights into the effects of pictorial health warnings.

Gospodinov and Irvine questioned the view that individuals are already well-informed about the consequences of smoking and therefore do not require additional information. Their concluding comments are at odds with the tobacco industry's argument that changes in information content or message format are unlikely to affect awareness or knowledge of the risks associated with smoking.

In summary, Gospodinov and Irvine's work identified a decline in consumption and a decline in prevalence; both findings are consistent with the hypothesis that the introduction of pictorial warning labels has affected smokers' behaviour. Todd's and Dixon's conclusions that this work showed no evidence of a reduction in the quantity of cigarettes smoked or a decline in the prevalence of smoking are not correct.

Liefeld (1999)

Both Todd and Dixon stressed the distinction between Type 1 and Type 2 research (or between the investigation of behavioural and psychological variables) that Liefeld (drawing on Professor Norman

Bradburn's work) established. Liefeld made this distinction in the context of a discussion of external validity and he explicitly acknowledged the difficulty of conducting Type 1 research. Liefeld was concerned that researchers may rely on measures of consumers' affective or cognitive responses, which may or may not be predictive of their behaviour. This concern highlights the tension between external validity, which is concerned with ensuring that the research is conducted in a manner that replicates or simulates the context in which the behaviour of interest is likely to occur, and internal validity, which attempts to control all extraneous variables so that the effects of the test variables can be estimated.

Liefeld suggested two approaches could be taken to Type 1 research, which he noted could be conducted in a test market situation, where respondents are not aware they are participating in a research study. Alternatively, he noted Type 1 research could use an experimental setting that simulates the marketplace experience respondents are likely to have when they purchase a product. He went on to suggest conjoint experiments as an example of this latter approach. It is important to note that Liefeld did not state conjoint experiments and test marketing were the only means of conducting behavioural research. Other research approaches can also estimate behaviour and achieve high external validity, although discussion of a wider range of alternatives appears to be beyond the scope of the research Liefeld was commissioned to undertake.

Liefeld set out a careful explanation of his research objectives and the analytical procedure used to analyse the data collected. His research investigated the relative impact of larger health warnings, stronger emotional and information content of health warnings; the effect of pictures, and the absence of trademarks and other identifying brand livery on encouraging cessation and non-initiation among smokers and non-smokers, respectively. Use of a conjoint analysis experiment enabled Liefeld to estimate the relative contribution of each attribute tested, and the importance of the different levels of attribute.

Liefeld asked smokers to indicate which of two packs they felt would most encourage them to quit smoking; respondents then used a scale to indicate how much more encouraging they felt the pack they had chosen to be. This process was repeated a further 11 times, until each respondent had evaluated 12 pairs of packs; the computer software used in the experiment selected the test pairs based on respondents' past choices. In addition, respondents were asked to rank 7 influences on their decision to stop or not start smoking; this question was administered twice (prior to exposure to the pictorial warnings and after respondents had seen the warnings). Respondents were also asked to identify the brand they were most familiar with from a montage of 29 brand packs that each had a health warning covering 35% of the front of the pack. They then repeated this exercise except that the packs now featured a warning message that covered 60% of the pack front. Details of respondents' demographics and smoking behaviour were also collected.

Liefeld concluded that pictorial health warnings were 60 times more encouraging than messages delivered in a text only format. He estimated that the message content accounted for 50% of the effect observed; the picture for 29%, the warning message for 12%, and the trademark for 8%. Liefeld's approach appears to have estimated the main effects of the model and he did not report any interaction effects between the attributes tested. This may be a function of the software he used, but details of the message-picture interaction would have provided further information about the effect of the graphic. Liefeld specifically noted that the relative contribution of the picture is under-represented, and stressed that warnings with no picture had no utility relative to those that did include a picture.

Todd did not criticise Liefeld's use of conjoint analysis. She suggested this research design allowed respondents to select from alternatives, removed problems of social desirability error that she alleged affected other studies, and enabled clear links to be made with respondents' likely behaviour. However, while she accepted Liefeld's research design and methodology, Todd emphasised his finding that the most important attribute was the message (or the presence of emotional information). She also emphasised his finding that the presence of trademark information was associated with encouraging smokers to quit (or non-smokers not to start). In both cases, she overlooked explanations Liefeld made with respect to these findings. He specifically stressed the important effect pictorial warnings had on encouraging smokers and non smokers alike. In addition, he offered an explanation for the apparently anomalous results relating to trademarks and a suggestion for future research that could test his explanation. Todd's failure to draw fully on Liefeld's wider explanation of his findings creates a misleading impression of his conclusions.

Unlike Todd, Dixon criticised Liefeld's research, which he argued does not estimate real behaviour. However, Dixon did not criticise Liefeld's suggestion that conjoint analysis met the requirements of Type 1 research nor did he demonstrate why an experiment that simulated marketplace decisions should lack external validity.

In summary, Liefeld's work represents a novel approach as it includes quasi-behavioural tests. However, asking respondents which image would most encourage them to quit, or remain a non-smoker, is not a measure of respondents' *actual* behaviour. Liefeld's work allows judgments to be made about the relative effectiveness of different attributes (and about the different levels at which these are expressed) on encouraging quitting or discouraging smoking initiation. However, while these estimates may provide insights into consumers' likely behaviour, they are not measures of actual behaviour.

Enviroics (2000)

This report outlined research that tested 67 messages and graphics in 35 focus groups conducted throughout Canada. The research comprised three phases, each of which tested a range of health warnings.

The research objectives included examining awareness of the health risks of smoking, and the current text-only warning labels shown in cigarette packages. In addition, the research explored participants' reactions to proposed pictorial warning labels, and their rating of specific health warnings. The research concluded that there was widespread awareness of the health risks of smoking and current health warnings, although specific estimates were not provided.

The vast majority of respondents thought cigarette packages should include health warnings; even "staunch smokers", who were less supportive of health warnings in other studies, agreed they were necessary. They felt these could remind smokers of the health risks of smoking, discourage smoking initiation, and help motivate those smokers contemplating quitting. However, the report noted that respondents were critical of the impact of health warnings, which some described as "drab and repetitive", and as having lost impact over time.

Respondents were shown a range of possible new health warnings and were asked to evaluate these. The report noted that most participants felt the inclusion of colour pictures represented an improvement over the existing warnings because of the images' ability to evoke more emotional responses and convey more information. Although some participants doubted the effect the new health warnings would have on smokers' behaviour, most felt the warnings could influence the behaviour of specific groups, particularly pregnant women, teens considering developing a smoking habit, and those attempting to quit. The remainder of the report provided detailed evaluations of each of the warning images tested.

EnviroNics (February 2001)

This research was conducted to provide baseline data prior to the introduction of pictorial health warning labels in Canada in January 2001. The study involved a telephone survey of the general Canadian population (N=721) conducted between November and December 2000. In addition, a further sample of smokers was interviewed during November and December 2000 to make a sample of 1000 adult smokers; a youth survey of 12 to 18 year olds (N=1015) was also conducted in December 2000. The response rates for these surveys were 28.2%, 17.5% and 20.8% respectively. The survey began by collecting details of respondents' smoking behaviour; 25% of adults smoked daily and a further 5% smoked occasionally, while 11% of youth smoked daily and 6% smoked occasionally. The researchers investigated cigarette consumption; respondents' main brand, and their attempts to quit smoking. They found that 77% of adult smokers had made an attempt to quit smoking and, of these, 60% had stopped smoking for at least 24 hours; 40% had made no attempt

to quit in the previous year. Sixty six percent of youth smokers had attempted to quit smoking and 83% of these had stopped smoking for at least 24 hours; 15% had not attempted to quit smoking in the previous year. Over half the adult smokers interviewed (56%) said they were seriously thinking of quitting; of these, 38% believed they would attempt to quit within the next 30 days and 45% within the next six months. An almost identical proportion of youth smokers (57%) indicated that they were also seriously thinking of quitting; 56% believed they would attempt to quit within the next 30 days and 28% would do so within the next six months.

Eighty seven percent of the adult sample thought cigarette smoking was a major health problem in Canada and a further 10% thought it was a minor problem. However, only 72% of smokers believed it was a major problem (cf. 91% of non-smokers), although a further 21% thought it was a minor problem (cf. 7% of non-smokers). Eighty four percent of youth thought smoking was a major problem while 13% thought it was a minor problem.

Top of mind awareness of the health effects of cigarette smoking ranged from 57% for lung cancer to just 1% for effects on unborn children. Adult smokers (49%) were less likely than non-smoking adults (59%) to associate smoking with lung cancer. This pattern also emerged when respondents were asked whether they agreed or disagreed with statements linking smoking to a range of health conditions. Although 96% of adult non-smokers strongly agreed that smoking caused lung cancer and a further 3% somewhat agreed, only 82% of smokers strongly agreed with this statement, with a further 12% stating they somewhat agreed with it. This pattern was evident across the range of health conditions mentioned. For example, 87% of adult non-smokers strongly agreed that smoking caused throat cancer (10% somewhat agreed) while 71% of smokers strongly agreed (22% somewhat agreed). Seventy eight percent of adult non-smokers strongly agreed that smoking causes heart disease (17% somewhat agreed) compared to 65% of adult smokers (24% somewhat agreed). In general, youth responses were similar to, or slightly lower than, those of adult non-smokers, with a few notable exceptions.

The survey also examined the effects of second hand smoke and smoking constituents; again, adult non-smokers and youth showed generally higher and stronger levels of agreement with the relationships tested than smokers. General population sample respondents identified nicotine (73%), tar (53%) and carbon monoxide (14%) as toxic substances in cigarettes; higher proportions of smokers than non-smokers or youth identified these substances, particularly tar and carbon monoxide.

General population respondents had seen information about the health effects of smoking in a variety of media of which television (62%) and newspapers (25%) were most frequently mentioned overall. However, 33% of smokers mentioned cigarette packages as an information source (cf. 20% of the

general sample and 13% of the youth sample); this information source was ranked second among those mentioned by smokers (after television).

Awareness of the health warning messages was very high (98% among smokers) and 51% of smokers indicated that they looked at health warning messages at least once a day. Among the general population sample, the most easily recalled feature of the health warning messages was the fact that they were words (53%) that featured in black and white (47%). Two thirds of smokers suggested the colour of the message while 55% noted that the message was text-based. When asked about specific health warnings, 58% of smokers recalled "smoking during pregnancy can harm your baby"; no other warning was recalled by more than 50% of smokers.

Smokers were also asked whether the health warning messages had affected them. Twenty one percent of adult smokers said the warning messages were very effective at informing them about the health effects of cigarette smoking, and a further 30% said they were somewhat effective; corresponding figures for youth smokers were 30% and 37%. Forty four percent of adult smokers said the warnings had been very effective or effective in getting them to smoke less around others over the last year (39% youth smokers), although 54% of adult smokers said they had not been very effective or were not at all effective in achieving this behaviour (61% youth smokers). Although 34% of adult smokers and 42% of youth smokers said the warnings had been very or somewhat effective in increasing their desire to quit over the previous year, 66% of adult smokers and 55% of youth smokers indicated the warnings had not been very effective or were not at all effective in this respect. Twenty seven percent of adult smokers (37% of youth smokers) said the warnings had been effective in getting them to smoke less than the previous year, but 72% of adult smokers (62% youth) said they had not prompted them to smoke less. While 26% of adult smokers said the warnings had been effective in prompting them to try to quit smoking, 74% said the warnings had not assisted them to try to quit (33% of youth smokers indicated the warnings had been effective while 66% said they had not been). Respondents agreed that health warnings provided them with accurate and important information about the health risks of smoking, although the levels of agreement were lower among smokers than non-smokers and youth.

Around a third of respondents was aware that new health warnings might appear on cigarette packages in the near future; awareness was higher among smokers than non-smokers or youth. Of those who were aware some changes might occur, the most frequently recalled change was the inclusion of pictures in the health warnings; again more smokers (72%) than non-smokers (55%) recalled this change, although 73% of youth mentioned this factor. When asked about their support for a proposal that cigarette packages include more detailed information about the toxic substances in cigarettes, 91% of non-smokers supported this measure, while 79% of smokers and 80% of youth supported it.

Environics (October 2001)

This report presented detailed results from a telephone survey of 2031 Canadian residents. The sample was stratified by geographic region and used quotas to ensure appropriate representation of specific demographic groups. The survey achieved a response rate of 14%.

The research found that 62% of respondents had noticed the changes to the warnings that appear on cigarette packages; 91% of daily smokers and 87% of occasional smokers had observed these changes. More specifically, 83% of smokers had noticed the incorporation of pictures into the health warnings; 19% had noticed that the size of the warnings had increased; and 10% noted that the messages themselves had become more strongly worded.

Twenty four percent of smokers had engaged in behaviours that reduced their exposure to the new warning messages (such as placing a sleeve over the cigarette pack or transferring the cigarettes to another container). Thirty percent of smokers who had concealed a warning message said they did so all the time and a further 37% did so occasionally (combined, these groups represent 16% of smokers in the sample).

When asked whether their knowledge of the health effects of smoking had increased or decreased since the introduction of the new health warnings, 14% of smokers responded that they now knew a lot more about health risks than they did before, while 21% of smokers considered they knew a little more. Fourteen percent of smokers said the new warnings had made them a lot more concerned about the health effects of smoking and 29% of smokers said they were now a little more concerned about these. Over 50% of smokers indicated they had frequently or sometimes discussed the new warnings with other people and that they thought more about the health effects of smoking. Fifty eight percent of smokers said the new warnings had made them think more about the health effects of smoking and 62% said the new warnings made cigarette packets look less attractive. Twenty one percent of smokers said the new warnings had prompted them not to have a cigarette when they were tempted to on more than one occasion, and 27% said the new labels motivated them to smoke less inside the home. Forty four percent of smokers said the new warnings had increased their motivation to quit, while 38% said the new warnings had been a factor in attempting to quit. Nearly half of the non-smokers interviewed said the new health warnings reinforced their decision to be a non-smoker.

Kaiserman, Illing & Dasko (2003) [Slide presentation]

This research was available only in the form of a presentation delivered to a conference in Helsinki in August 2003. The presentation described the change in health warning messages from text-only to pictorial warnings and noted that the purpose of this change was to inform consumers about the

hazards of tobacco use; encourage cessation; reduce consumption, and encourage smokers to avoid consuming tobacco products where this could harm others. The presentation referred to the baseline survey (wave 1) described above (Environics February 2001) and a series of subsequent surveys conducted 3, 6, 12 and 18 months after the introduction of pictorial warning labels in Canada. Sample sizes were approximately 1500 (1000 non-smokers and 500 smokers); details of the response rates achieved were not provided.

The survey results showed some evidence of a decline in smoking behaviour among adults (30% to 28%) and youth (17% to 13%), and a decline in the proportion of potential youth smoker (36% to 32%). However, no tests of significance were provided and the data contained fluctuations from one wave to the next, making it difficult to identify clear trends.

The proportion of adult and youth respondents who reported attempting to quit once or twice increased from the baseline survey to the wave 5 (18 months) survey. The proportion who attempted to quit three times appears to have declined from the baseline (10%) to 18 months (6%) following the introduction of the new warnings, though this is most likely due to the increase in the proportion who had attempted to quit on more than three occasions. The proportion that had made no attempt to quit in the last year declined sharply from wave 1 to wave 5 for both adult and youth smokers.

Although television remained a key medium for providing information about the health effects of smoking, cigarette packages had clearly become a more important medium for adult and youth smokers and potential youth smokers, and cigarette packages were the most frequently mentioned information source by youth smokers. Overall, the proportion of adult and youth smokers mentioning cigarette packages as an information source was nearly double that recorded in the baseline survey (33% of adults to 58% in wave 5; 34% of youth to 60% in wave 5; 14% of potential youth smokers to 37%). The presentation also included results showing an increase in top of mind awareness of health risks associated with smoking, and higher levels of agreement that smoking resulted in a range of health conditions.

The proportion of respondents who looked at the health messages showed fewer adult, youth and potential youth smokers looked at the health warning messages several times a day in wave 5 than in the baseline survey, just before the messages were introduced. The proportion of potential smokers who looked at the messages once a day or every 2 to 3 days had increased, as had the proportion of adult and youth smokers who read or looked at the messages about once a week. The proportion of all groups who stated they never looked at or read the warning messages increased from the baseline to wave 5.

Recall of specific on pack warning messages increased among adult smokers from wave 2 (3 months after the introduction of pictorial warning labels) to wave 5 (18 months following introduction), but was more variable among youth smokers, where the levels fluctuated, but generally stayed at similar points across the waves. The proportion of respondents who thought the health warning messages had been effective at informing them of health risks associated with smoking increased among both adult smokers (54% in wave 2 to 68% in wave 5) and youth smokers (74% in wave 2 to 83% in wave 5). Adult smokers were more likely to see the packs as effective in prompting them to quit (36% in wave 2 to 45% in wave 5) than youth smokers (50% in wave 2 to 44% in wave 5, although prior to wave 5, youth smokers' responses had followed the same pattern as adult smokers). The proportion of adult smokers, youth smokers and potential youth smokers evaluating the messages as accurate and providing important information remained steady at high levels (80%+ for all groups).

Overall, Kaiserman, Illing and Dasko concluded that smokers' awareness of the health risks of smoking remained high, as did their support for the new messages. They also concluded that smokers continued to notice and read the messages, that their learning had increased, and that the messages did not show signs of wearout after 18 months.

Todd argued the first Environics study indicated that respondents were more likely to use television and newspapers, rather than cigarette packages, as sources of health information. However, the results she appears to be drawing on relate to the general adult population estimates; 33% of adult smokers indicated they had seen health information on cigarette packages (the second highest ranked medium) and 13% of youth had seen health information on cigarette packages (the fourth highest ranked information source, although the estimate is based on the wider youth sample and not youth smokers).

Todd also argued that the Environics (February 2001) study showed strong agreement with the health statements examined in the research. Although more than 50% of smokers surveyed strongly agreed that smoking could cause lung, throat and mouth cancer, heart disease, emphysema, asthma, and chronic bronchitis, less than 50% strongly agreed that it would cause premature death, reduced growth of babies during pregnancy, gum disease, premature ageing, stroke, premature birth, or blood clots. Total agreement that smoking could cause stomach ulcers, miscarriages, impotence, bladder cancer, infertility, gangrene, Alzheimer's, acne, MS, arthritis or hepatitis was less than 50% among smokers. Less than 50% of adult smokers recalled the specific health warning messages that feature on cigarette packages, with the exception of "smoking during pregnancy can harm your baby", which 58% recalled. Closer examination of the research findings suggests smokers' knowledge of the wider health consequences of smoking cannot accurately be described as comprehensive; their ability to recall specific health messages was also highly variable.

Todd used the information relating to the impact of the existing messages selectively. Forty four percent of adult smokers said the messages had been effective in prompting them to smoke less around others over the last year (although 54% said they had not been effective). Over a quarter (27%) said the messages had prompted them to smoke less (cf. 72% for whom the messages were ineffective), or had prompted them to attempt to quit (26% cf. 74% for whom the messages had been ineffective). The proportions of youth smokers who found the messages to be effective in prompting them to reduce smoking or attempt quitting was generally slightly higher than recorded in the adult smoker sample. These results suggest that a minority of smokers indicated they were responsive to the text-only warning labels, but the majority declared themselves to be unmoved by these warnings.

Todd's comments on the Environics (October 2001) study, which she mistakenly stated is the earlier of the 2001 reports, are confined to methodological comments. In particular, she comments on the low response rate and the potential for non-response error. The response rate is not 8% as Todd states, but 14%. Notwithstanding Todd's error, the response rate is low by any standards and low response rates introduce the possibility that people not able to be contacted, or not willing to participate in the research, may differ from those whose opinions and behaviours were recorded. Non-response error is not, as Todd implied, inevitable; surveys with low response rates may be unaffected by non-response error, just as estimates from surveys with high response rates may be affected by this error. However, because smoking is correlated with socio-economic status, and the data were weighted to ensure employment status, income and occupation corresponded to the Census data, the potential for non-response error would seem to be reduced. (There is no *a priori* reason why blue collar workers who could not be contacted should hold markedly different views from those who were successfully interviewed.) Nevertheless, we are not able to reject the possibility that non-response error affected these estimates.

Todd also criticised Kaiserman, Illing & Dasko's research, which she argued indicated strong awareness of health warnings, and strong support for the content of these, but failed to demonstrate a link between recall and behaviour change. Overall, she concluded their results showed the previously high levels of awareness had been maintained, but that there was no evidence awareness had resulted in behavioural or attitudinal change.

Again, Todd's comments seem at odds with some of the evidence presented in Kaiserman, Illing & Dasko's presentation. This suggested that general declines in adult, youth and potential youth smokers had occurred, although there were fluctuations in the data presented. However, the proportion of adult and youth smokers who had attempted to quit once or twice in the last year had increased; although the proportion who had attempted to quit three times had decreased between wave 1 and 5, it is logical to attribute this to the increase in the proportion who had attempted to quit

on more than three occasions. Furthermore, the proportion that had not made any attempt to quit showed a sharp decline in both groups between waves 1 and 5.

Todd did not comment on the large increase in all groups (adult and youth smokers, and potential youth smokers) who cited cigarette packages as a source of health information. Contrary to Todd's assertion that awareness remained at existing high levels, it is clear from earlier research that awareness of several smoking-related health effects was not high. Kaiserman, Illing & Dasko's presentation showed an increase in awareness of health messages; they also showed an increase in agreement that smoking causes a range of adverse health effects.

Comparison of results with the February (2001) Environics report, reveals that the proportion of smokers who felt the health warning messages had been effective in prompting them to try to quit smoking had increased. In the February (2001) report (which was based on data collected prior to the introduction of pictorial health warnings) 26% of adult smokers said the warnings had been effective in prompting them to try to quit smoking, and 33% of youth smokers indicated the warnings had affected them in this way. Kaiserman, Illing & Dasko's presentation showed that the proportion of adult smokers who said the health warnings had prompted them to attempt to quit increased to 45% 18 months after the new warnings had been introduced (after showing steady increases over each survey wave). The proportion of youth smokers who said the health warnings had been effective in prompting them to attempt to quit increased to 55% 12 months after the pictorial warnings were introduced, although this figure reduced to 44% in the wave 5 survey. Even using the lower figure, it is clear that substantially higher proportions of smokers said the new health warnings had been effective in prompting them to attempt to quit.

Dixon criticised the MoH consultation document's use of the Environics research, which he argued had been mis-quoted. He noted that the Environics (February 2001) survey reported baseline results against which the effects of the new health warning messages could be assessed whereas the Environics (October 2001) survey was not a follow-up survey, but a separate study. Dixon is correct in noting that the October 2001 survey was not an explicit follow-up survey; however, the sample selection procedure and survey mode (telephone) appear to be identical in both surveys. Because many of the questions differed across the two surveys, it is difficult to draw explicit comparisons between them since even quite small changes in question wording can alter the response distribution.

The Environics (February 2001) survey asked respondents how effective on-pack health warning messages had been in prompting them to perform a range of behaviours. Of the respondents to the earlier study (February 2001), 21% of the adult smokers said the warnings had been more effective in prompting them to smoke less around others (23% said they were somewhat effective). The Environics (October 2001) survey asked respondents whether the on-pack warnings had prompted

them to smoke less inside their home. It is clear that this question differs from the earlier question and it is thus not surprising that the estimates also differ. Eleven percent of smokers reported smoking much less inside their home; 16% smoked somewhat less, while 72% said the new health warnings had had no impact on how much they smoked inside their home. The most obvious explanation for this apparent discrepancy is that many people may not see smoking inside a home as inconsistent with not smoking around others (for example, people may smoke in a bedroom, conservatory, study or other room not used by other household members).

The second question that contained similar idea elements related to quitting attempts. The February (2001) survey asked respondents how effective the health warnings messages had been in getting them to try to quit smoking within the past year, while the October (2001) survey asked respondents to what extent the new warnings had increased their motivation to quit smoking. Twenty six percent of respondents to the first survey indicated that the warnings had been effective in getting them to attempt quitting (9% said they were very effective and 17% somewhat effective; 74% said they were not effective). By contrast, forty four percent of the respondents to the October survey indicated that their total motivation to quit had been increased by the new warning labels (11% by a lot; 33% by a little). More importantly, of the occasional smokers, total motivation to quit had increased to 60% (unfortunately the earlier study did not present analyses broken down by frequency of smoking).

Although Dixon argues that the first survey cannot be treated as a baseline against which results from the second can be compared, we believe the questions exploring quitting draw on the same concept and allow comparisons. The estimates from the second survey suggest the introduction of pictorial warning labels had increased smokers' motivation to quit, and we believe the question used in the February (2001) survey is sufficiently similar to the October (2001) question to be used as the basis of comparison.¹ However, estimates of motivation to quit are not measures of smokers' actual behaviour, although changes in motivation are often assumed to precede behaviour.

Dixon also claimed that Health Canada had conducted several follow-up studies to the February (2001) baseline, but had not placed these in the public domain. Although Dixon has apparently accessed details of these studies using the Freedom of Information Act, he did not present any results from these and instead argued that they have not been associated with a reduction in cigarette consumption, an increase in quitting attempts, or an increase in the proportion of smokers believing smoking is a serious health problem. Without details of the actual data, it is impossible to assess Dixon's conclusions.

¹ Labaw (1980) noted that question wording variations become significant where they introduce or tap a different concept or relating or emotional level surrounding an issue.

Dixon also claimed that Kaiserman *et al's* (2003) presentation avoided mentioning the alleged lack of effect graphic health warnings have had on key smoking indices. Our examination of Kaiserman *et al's* (2003) presentation suggests the research found increases in the proportion of smokers attempting to quit (and reductions in those who had not attempted to quit). Higher proportions of both adult and youth smokers stated that the information on cigarette packages had been important in prompting them to attempt to quit. Respondents' use of cigarette packages as a source of information about the health effects of smoking increased, as did agreement that smoking could cause a range of medical conditions. These findings are at odds with Dixon's conclusions.

Overall, although both Todd and Dixon criticised the Environics studies, a number of the concerns they raise can be addressed by a broader examination of the data. The studies, whether part of a broader programme or ad hoc pieces of research, support the conclusion that the pictorial health warnings introduced in Canada in January 2001 have increased smokers' knowledge of the health risks associated with smoking, and smokers see the pictorial warnings as more effective in prompting them to engage in cessation related behaviours. However, while the data contain some measures of self-reported behaviour, much of the data is perceptual and cannot be assumed to be equivalent, or causally related, to behaviour.

Hammond, Fong, McDonald, Cameron & Brown (2003)

The research outlined in this paper examined the impact of the Canadian pictorial warning labels on adult smokers and involved two phases: an initial survey conducted in late 2001 that was followed up by a second survey three months later. The initial survey (N=616) investigated respondents' smoking behaviour and demographics, and their knowledge of the warning labels, and explored the extent to which they processed the information in these. The response rate was 76%. The extent of cognitive processing was measured using a number of variables that examined how carefully respondents read the on-pack labels, thought about these, and discussed them with others. The survey presented respondents with nine items that they responded to using a 5-point scale ranging from "not at all" to "all the time". The follow-up survey (N= 432) examined the same variables and asked respondents if they had made any changes in their smoking behaviour, such as attempting to quit or reducing the number of cigarettes they smoked.

Hammond *et al* reported that respondents had a strong awareness of the warning labels; over 70% correctly recognised three actual and three fictitious (control) messages. They also found that the level of cognitive processing was generally high; over 90% of respondents had read the on-pack messages and just under 80% reported having paid close attention to these. When responses to the two survey waves were compared, Hammond *et al* found that the proportion of respondents who had read, paid close attention to and thought about the messages on the inside of cigarette packages had significantly increased, while the proportion who had talked about the new warning labels with others

or who had thought about the warnings when a cigarette pack was not in sight had decreased significantly. They also reported that intention to quit was significantly related to cognitive processing.

The follow-up study revealed that 23% of respondents had attempt to quit in the interval between the two surveys and nearly 11% of these had not resumed smoking by the time of the second interview, and nearly a quarter of these had reduced the number of cigarettes they smoked each day. The depth of cognitive processing of the messages, prior attempts to quit and intention to quit were all significantly and positively associated with cessation behaviour, while the number of cigarettes smoked each day was significantly and negatively association with cessation behaviour.

Hammond *et al*/noted limitations to their research, particularly the lack of power, which affected their analyses and whether the relationships tested were significant. They also suggested future research could be conducted over a longer time period, which would allow more detailed exploration of cessation behaviours and examination of any wearout affecting the warnings.

Todd commended Hammond *et al* for their use of behavioural outcome variables and their development of an involvement index, which she viewed as a more sophisticated measure than estimates of awareness or knowledge. However, she argued that Hammond *et al*/could not attribute the changes they observed to the warning labels alone, as other interventions may also have affected respondents' behaviour. While it is correct to note that a wide range of factors may influence respondents, researchers cannot take all of these into consideration and so normally assume their results apply "other things being equal". Without direct evidence of external factors that may have affected the findings (thereby demonstrating that this assumption was violated), it is not possible to test Todd's concern.

Todd also noted that the study did not explain why some smokers engaged in more cognitive processing than others. The most likely explanation of this is that smokers' responsiveness to health warning messages is clearly related to the length of time they have been smoking. Canadian research (Environics, 2000) and Australian research (Elliott & Shanahan, 2000) documented older smokers' resistance to health warning messages and this is a likely explanation of the different levels of cognitive processing observed. However, although logical, this explanation needs to be tested against Hammond *et al*'s data before it can be accepted.

Dixon also criticised Hammond *et al*'s work, which he argued was somewhat misleading in its attribution of behaviour change to pictorial health warnings. Dixon argued the changes observed could be due to the use of health warnings in general; that is, he suggested that a lag effect from the earlier text-only warnings had not been considered. Dixon also alleged that the absence of data

collected prior to the introduction of pictorial health warnings meant that it is not possible to attribute the changes observed specifically to the introduction of graphic images in on-pack warnings. The absence of benchmark data means that the precise role played by the pictorial warnings is difficult to quantify. However, evidence from other studies supports Hammond *et al*'s conclusion that the pictorial warnings stimulated respondents' cognitive processing. The October (2001) Environics survey noted that 35% of respondents felt they knew more about the health effects of smoking since the introduction of the new warning labels; 41% were more concerned about these issues, and more than 50% of smokers had frequently or sometimes discussed the new warnings with other people. These data suggest the pictorial warning labels prompted an increase in smokers' perceived knowledge of the health effects of smoking and the salience of these.

Dixon also argued that the direction of causation proposed by Hammond *et al* was not clear, and that smokers who intend to quit may be more likely to read health warnings, rather than vice versa. It is not possible to determine the direction of causation from the data presented, although the alternative direction Dixon posited suggests some other factor has contributed to smokers' intention to quit and their greater responsiveness to the new health warnings. Dixon offered no evidence to support the direction of causality he proposed in the absence of this it is difficult to evaluate his argument further.

Finally, Dixon alleged that awareness of health warnings was high prior to the introduction of pictorial health warnings and that the use of graphics alone could have done little to increase the already high levels of awareness. While there is strong evidence that smokers are aware cigarette packages feature health warnings, their awareness of the specific warnings featured is much lower (see Environics, February 2001).

Hammond, Fong, McDonald, Brown & Cameron (2004)

This paper is based on the same data set analysed in this research team's 2003 paper, discussed immediately above. This study examined the criticism that graphic warnings cause unnecessary distress to smokers and prompt reactance behaviours that may ultimately result in an increase in consumption. The variables measured are the same as those discussed in Hammond *et al* (2003) and also included measures of avoidance behaviours, such as attempts to cover or hide the labels, or requests for specific packages to avoid particular warnings. An index of negative emotional reaction to the labels was developed on the basis of responses to fear and disgust questions, and the credibility of the warnings was also assessed.

Hammond *et al* reported that 63% of respondents indicated they had received at least one cessation benefit while only 6% reported a negative outcome. Thirty six percent of respondents had made some effort to avoid the labels, although avoidance was not associated with either depth of processing at the baseline or cessation behaviours at the follow-up. Forty four percent of

respondents reported experiencing fear after viewing the labels and 58% reported feeling disgust. Hammond *et al* found that respondents who reported experiencing higher levels of fear and disgust were more likely to have read and thought about the warning labels. Similarly, smokers who reported experiencing higher levels of fear and disgust were more likely to have reported quitting, attempting to quit, or having reduced the number of cigarettes they smoked. Although 13% of smokers felt the warnings did not report the health risks of smoking accurately, and 27% felt they contained too much information, 50% wanted cigarette packages to feature more information.

Although Hammond *et al* noted the lack of baseline measures collected when the text-only messages were used, and the absence of a control group, they argued that their findings were consistent with other studies. Furthermore, they concluded that claims graphic images led to adverse outcomes did not appear to have any empirical support.

Australian Research

Borland & Hill (1997)

This study assessed new health warnings and product labelling introduced in Australia in 1995, and explored awareness of the new warnings during the implementation period and changes in knowledge once the new warnings had been introduced. Two surveys, one in late 1994 (N= 1035) and the other in mid-1995 (N=1033) were conducted; the follow-up survey included 243 smokers from the first survey who had been successfully re-contacted and who agreed to participate in a second interview. The response rate was reported to be 66% to both surveys (though no specific calculations were provided). The questionnaire explored respondents' smoking behaviour, their knowledge of a range of variables including smoking related illnesses and the risks of smoking; their recall of the new warnings, and their views on these.

Borland & Hill reported that smokers' awareness of the new messages had increased from 28% at the baseline survey to 91% at the follow-up, while non-smokers' awareness increased from 24% to 51%. Respondents' knowledge of the health warnings was also high, although it seems likely that these estimates were influenced by earlier questions investigating their beliefs about the six new warning statements. Nevertheless, recall of warnings increased among the sample of re-interviewed smokers from a mean of 1.9 to a mean of 2.8 warnings recalled.

Smokers were slightly more likely to explain what tar was in the follow-up survey than in the baseline study, although there was no difference in their ability to explain the effect of tar on smokers. The proportion of smokers able to answer questions about nicotine content of cigarettes also increased across the surveys, as did the proportion of smokers able to give an acceptable response to a question asking them to explain what carbon monoxide was. The smokers in the re-contact sample showed no significant change in knowledge of tar, but an increase in knowledge of what nicotine and its effects are, and in knowledge of carbon monoxide and its effects.

Just over thirteen percent of smokers reported that the new warnings had led them to smoke less, although response to an earlier question that showed 19% claimed to have reduced consumption did not differ from the baseline estimate of 16%. Only one smoker in the entire sample had called the telephone information line promoted on packs, although another 5% of smokers had considered doing so.

Smokers in the follow-up survey were more likely to mention smoking related illnesses featured in on-pack warnings, although smokers were still generally less likely than non-smokers to mention these. Smokers were also less likely than non-smokers to agree with statements that linked illnesses mentioned in on-pack warnings with smoking. However, the follow-up survey showed an increase in the proportion of smokers that believed smoking was addictive and able to harm others. Borland &

Hill concluded that the new health warnings had improved knowledge of the risks of smoking and increased the salience of this knowledge.

Dixon did not comment in detail on Borland & Hill's work, presumably because their study investigated changes to text-based warnings, rather than the introduction of pictorial warning labels. However, although Todd accepted that smokers in the second survey were more knowledgeable than smokers interviewed in the first survey, she commented that non-smokers' knowledge had not increased in the same way. This would seem very likely to be a consequence of non-smokers' general lack of exposure to on-pack warnings. The critical point in Borland & Hill's research is that the salience of the health warnings had increased among smokers.

Todd also commented on the possibility of a testing effect among respondents, some of whom participated in the survey twice. Participation in an earlier study may have sensitised respondents to the questions asked and, without analyses that enable a comparison of estimates based on respondents who participated in the first and second surveys with those from respondents to the second survey only, the existence and extent of any testing effect cannot be established. In addition to this effect, it is likely that the question order also influenced respondents. Borland & Hill mention and discuss both potential problems, although they do not discuss the likely effects of these in detail.

Todd also asserted that smokers' agreement that smoking is addictive constitutes evidence that smoking is not based on a rational decision and is therefore unlikely to be influenced by increases in knowledge. The corollary of this reasoning is that smokers will be more responsive to emotional arguments, an argument advanced by supporters of pictorial warnings, which several qualitative studies have established evoke much stronger emotional responses than text-only warnings.

Elliott & Shanahan (2000)

This project involved several phases: a literature review, a quantitative study of 1204 Australians aged 15 years and over (response rate not provided), and a series of interviews with stakeholders. The literature review identified several variables thought to affect the effectiveness of health warnings and responses to these. Data from the quantitative study were weighted to correspond with Australian population data; key estimates were compared to other studies and were generally in keeping with the estimates reported in these. The quantitative study revealed that younger smokers were less concerned about the health risks of smoking than older smokers, who tended to rationalise their behaviour in a variety of ways, including rejecting information about the health-related consequences of smoking.

Respondents regarded the health warnings on cigarette packages as important and 71% of current smokers believed they were very or quite important (by contrast 86% of non-smokers considered

them important or quite important). These figures were similar to those found in an earlier benchmark study (Elliott & Shanahan, 1996), although the proportion of smokers describing the warnings as “very important” increased from 43% to 49%. Awareness of the front label information on cigarette packages also remained high and virtually all recent ex-smokers and smokers were aware that health warnings featured on the front of cigarette packages, and over 90% of these groups had read the information. Although these figures remained constant across the 2000 and 1996 surveys, respondents in the 2000 survey had lower awareness of the information on the side and back of cigarette packages.

Unaided recall of the health messages varied; although 65% of smokers recalled “smoking when pregnant harms your baby” and 50% recalled “smoking causes lung cancer”, recall levels of other health messages were all below 50%. Predictably, aided recall levels were much higher and recall of specific warning messages was similar across both surveys, with the exception of “your smoking can harm others”, which 86% of respondents recalled in 2000 (cf. 71% in 1996). Smokers’ knowledge of the constituents of cigarettes, and the effects of these, had generally increased in 2000 relative to 1996, although in all cases substantial groups were unaware of the tar, nicotine and carbon monoxide content of their preferred brands, and over a quarter did not know the health effects of these ingredients.

The 2000 survey explored smokers’ views on the importance of health warnings; there was a significant increase in the proportion that said these were very important (43% in 1996 cf. 49% in 2000). Smokers’ attitudes toward the health warning labels were similar across both surveys; however, Elliott & Shanahan noted that higher proportions of smokers and recent non-smokers agreed that “The health warnings on cigarette packs should be stronger”. The proportion of smokers whose knowledge of the health effects of tobacco had been improved a lot from the inclusion of health warnings and information on cigarette packs increased from 29% to 32%, although the proportion who indicated the warnings made no difference remained at 40%. The health warnings’ effect on smokers’ decision to smoke less, switch to a lower tar brand, or quit remained similar from 1996 to 2000. However, the proportion who said they had recently reduced the amount they smoked each day had increased from 24% to 29%, even though the proportion that had done nothing increased slightly to 51%. This latter figure showed some difference across gender, as the proportion of male smokers that reported doing nothing different decreased from 55% to 50%, while the proportion of females that made the same claim increased from 44% to 52%.

When discussing factors that aided their decision to quit, 16% of recent ex-smokers noted health warnings had played a role in their decision (this factor ranked 7th out of the seven factors examined). The most important factor was that smoking was affecting their health (64%) followed by costing them too much (47%). Twenty six percent noted that health warnings on television had

influenced their decision. The survey also examined smokers' attempts to quit and their awareness of an information line (60% were aware, an increase of 20% over the previous survey).

Stakeholders strongly supported changes to the current health warnings that would strengthen these and they made a number of suggestions about how the text-only warnings could, in their opinion, be improved.

Elliott & Shanahan (2002) Stage 1 Research

This research comprised forty four mini group discussions (each involving four or five participants). These were undertaken with current smokers, recent and longer-term ex-smokers, and non-smokers. The research involved testing a wide range of potential health warnings and evaluating these with respect to their noticeability, comprehensibility, believability, memorability, informativeness, persuasiveness and visibility. The research examined responses from young smokers (15-17; 18-24); mature adult smokers (25-49) and older smokers (50-70). Elliott & Shanahan observed that the warnings tested evoked differing responses from these groups. They found young male smokers were less inclined to see the messages as personally relevant, since they viewed the health consequences of smoking as something that affected only older smokers. Moreover, since a number indicated they planned to quit smoking in their 20s, they felt confident the health consequences described or depicted would not affect them. However, female smokers in the 15-17 age group were responsive to the graphic images they viewed, particularly those depicting health problems and the effect of smoking on their appearance.

Elliott & Shanahan found that although smokers in the 18-24 age group admitted they were addicted to smoking, many nevertheless thought they would not suffer smoking-related health consequences because they would quit before these effects became likely to affect them. However, some issues, particularly those relating to skin ageing and pregnancy affected this group and the graphics affected females and males alike.

Smokers aged 25-49 had varied life experiences and these shaped their responses to the messages and images. Those with children, or contemplating families, were more affected by messages about the effects of smoking on children or while pregnant. Respondents in their 20s and 30s were less concerned about the health effects of smoking while those in their 40s were more likely to know someone who had suffered health problems attributed to smoking or to have suffered these problems themselves, and they were more responsive to the health messages. The graphics had a generally high impact on this age group and affected both genders. Images depicting easily identified body parts were more effective than images of internal body organs, some of which were less easily recognisable.

Older smokers displayed more entrenched attitudes and behaviour and many had smoked for several years. Because of their long history of smoking, a number thought their health had already been damaged and that quitting would not alleviate this damage. This group was the least positive about the health messages and rejected those they felt were at odds with their own experience, although they were more likely to accept warnings based on medical facts viewed as common knowledge. The group felt it was important that young people were warned about the addictive nature of smoking. Despite their ingrained attitudes and behaviour, the graphics also affected this group and respondents felt they would be hard to ignore and that they could prove an effective deterrent to new smokers and those contemplating smoking.

Non smokers and ex smokers were most supportive of health warnings, which they viewed as reinforcing the growing social unacceptability of smoking. Elliott & Shanahan provided details of respondents' reactions to each of the health warnings and graphics used in the discussion groups. They concluded that health warnings would help increase the social unacceptability of smoking, generate controversy and discussion, and that they were likely to assist quitting, rather than being the sole motivating factor that prompts quitting attempts. In addition, Elliott & Shanahan suggested characteristics the research indicated would improve the effectiveness of health warnings. These included rotation of the warnings to maintain "freshness", links with television campaigns, use of clear and easily recognisable visuals and brief simple messages; use of "new evidence", increase in the font size and overall warning size to improve readability, and inclusion of a "way out" for smokers.

Elliott & Shanahan (2003) Stage 2 Research

This research was designed to follow the study summarised above. It used a similar methodology (twenty mini-group discussions conducted with a sample of smokers and non-smokers) and addressed similar objectives, although fewer messages were tested and those examined were reviewed in greater detail. This particular work also used mock-up packs featuring pictorial warnings to introduce the specific stimulus material to respondents and to increase the external validity of the research.

As with their earlier research Elliott & Shanahan found that entrenched smokers were most resistant to the new health warnings, including the graphics, while those smokers contemplating quitting responded most favourably to them. Elliott & Shanahan reported that the graphic warnings evoked the most dramatic response from smokers, who described them as more impactful, more attention-getting, and harder to deflect. Smokers also considered the graphics were more confronting and thought they would act as a strong deterrent to new smokers or those considering developing a smoking habit. They concluded that the packs featuring text only warnings were not as impactful or effective as the packs featuring graphics at conveying the negative health consequences of smoking.

Elliott & Shanahan noted that graphics depicting external visual effects of smoking were more arresting and memorable, while less effective graphics either featured images that participants found difficult to recognise or they did not have a clear link to a specific health issue. They also reported that respondents thought too few of the messages tested had a positive orientation, that respondents did not believe some of the messages, particularly where they did not link the diseases shown with smoking. Elliott & Shanahan recommended the development of new images and messages to address the issues raised by participants in the group discussions.

Overall, Elliott & Shanahan concluded that inclusion of graphics in on-pack health warnings would heighten the salience of the health warnings, but could risk raising anxiety among some smokers who see themselves as purchasing a legal product and engaging in a legal behaviour. Nevertheless, they concluded that the graphics made the health warnings more noticeable and suggested that people seeing these would be less likely to deflect the health warnings associated with them. They also concluded that inclusion of graphics would be likely to negate the pleasure of smoking and make smokers feel uncomfortable about their habit. However, they noted that the reaction elicited would depend on the particular "state of change" that best applied to the individual smoker. While they found evidence that the graphics may reduce smoking initiation and support quitting attempts, they also noted that they may make hardened smokers more defensive.

Both Todd and Dixon criticised Elliott & Shanahan's work; because their criticisms tend to apply to all three studies discussed above, we consider their comments on the three studies jointly. In discussing Elliott & Shanahan (2000), Todd noted that the research enabled some comparisons to be drawn, although it was not truly longitudinal as the same respondents had not been interviewed in both the 1996 and 2000 studies. She noted that the findings from the 1996 and 2000 surveys are largely similar, that the proportions calling the information line had not changed, even though awareness of the information line increased. She argued that the study's focus on awareness and attitudinal measures was problematic.

Dixon was more critical of this research, which he argued provided no evidence to support the argument that text-only warnings had become less noticeable over time. In particular, he claimed the research did not show that awareness of warning labels or the content of these had declined between the two surveys, and he suggested there was some evidence that it had increased. However, this argument does not directly address the conclusion that the text-only labels had become less impactful and suffered from wearout. It is quite consistent for awareness to remain high (respondents recall information they have learned) and yet for that information to have lost salience.

An alternative view of the data suggests that although 71% of smokers considered the health warnings important, they were not important for over a quarter of smokers (Elliott & Shanahan, October 2000). Although aided recall of a number of the text-only messages used was high, this does not mean those messages are highly salient to smokers, or that changes to the format and content of health warnings would not increase the warning salience. Similarly, although smokers reported that their knowledge of the health effects of smoking had improved a lot because of the health warnings, 40% stated the warnings had made no difference to them, 20% said the warnings had not affected their behaviour, and, when asked about their recent smoking behaviour, 51% said they had done nothing different. While smokers' level of awareness remained similar across both surveys, it is also clear that the proportion of smokers unresponsive to the text-only warning messages also remained similar across both surveys. This finding appears to support Elliott & Shanahan's argument that changes to the format and content of the warnings may decrease the proportion of smokers who describe themselves as unaffected by the warnings. In addition, the proportion of smokers who agreed that the on-pack warnings should be stronger had increased. Together these findings indicate that smokers may be suggesting they would be more responsive to stronger warnings, a conclusion consistent with their increased perception of the risk they face from smoking.

Todd argued that the definition of "effectiveness" used in the 2002 and 2003 research was vague and that it appeared to correspond to the amount of information conveyed rather than to a specific smoking (or non-smoking) behaviour. Elliott & Shanahan did define the measures of effectiveness their research tested (these included noticeability, comprehensibility, believability, memorability, informativeness, persuasiveness and visibility of the messages and graphics). These measures all relate to the communication effectiveness of the messages and graphics, and they do not necessarily provide insights into smokers' likely response to the health warnings tested. Elliott & Shanahan did include some proxy variables that are often assumed to provide insights into behaviour; for example, they investigated the graphics and text most likely to trigger a response to reduce tobacco consumption or a quitting attempt. Similarly, they explored the persuasiveness of the graphics and the types of reactance behaviours that might be associated with the packs. However, these variables are not measures of actual behaviour and while the research enables a comparison of the relative performance of the graphics and messages on these variables, it does not enable an estimation of the effect these will have on the behaviour of smokers in general.

Todd also argued that the group discussion protocol was very structured and that there was little evidence other issues had been raised. In our experience, group discussion protocols are often structured in this way to ensure the discussion covers a wide range of topics and that these are explored in detail. We have never found use of a protocol to inhibit participants from raising other issues. Similarly, it is not unusual to ask respondents to complete a short questionnaire before a

group discussion commences. So long as the questions used are not leading, there is no reason to suppose that this practice would sway respondents' opinions in a particular direction or introduce bias to the subsequent discussion. In this case, the questionnaire involved a comparison of the health information conveyed by text-only packs and text-graphic packs. This questionnaire would clearly have introduced respondents to the discussion topic, but Todd provides no evidence that respondents' opinions were led in any way.

Without access to full transcripts of the group discussions, it is not possible to comment on Todd's assertion that other issues were not raised during the course of the discussion, nor is it possible to comment on whether the moderators influenced the tone of the discussion. The wide range of views elicited, and the fact that "entrenched" smokers clearly felt able to make adverse comments about the health warnings, suggest that respondents were in an environment where they felt able to express their views, whatever these might be. While social desirability error is difficult to discount, the diversity of views held, together with the fact that the groups comprised either smokers or non-smokers, suggests the research environment facilitated open expression of respondents' opinions.

Todd also asserted that the public sharing of information on a socially sensitive topic may have affected the findings obtained, although she did not suggest what specific effects concerned her. Again, the range of views elicited would suggest that respondents felt able to express varying opinions. However, without the discussion transcripts, it is not possible to comment on the group dynamics or the extent to which the group composition affected the tone and direction of the discussion, if in fact it affected the discussion at all.

Todd questioned the "representativeness" of the groups used in the research. We believe this may mis-interpret Elliott & Shanahan's statement about the group participants, whom they noted represented "a range of culturally and linguistically diverse backgrounds and a range of socio-economic strata" (p. 19). This is not a statement that the research groups are representative of these population sub-groups, but rather a statement that these sub-groups are represented within the discussion groups.

Dixon also criticised Elliott & Shanahan's qualitative research, which he argued is less reliable because the samples sizes are smaller than those used in quantitative research. Where the samples are convenience samples rather than probability samples, sampling theory does not apply. It is not possible to generalise from qualitative research based on convenience samples with any known degree of precision or accuracy. However, Elliott & Shanahan did not attempt to estimate population parameters and they describe their research as developmental; this suggests they were aware of the limitations affecting their data.

Dixon also argued that evidence of the alleged wearout of text-only health warnings comes not from the group discussion participants, but from the stakeholders interviewed as part of the 2000 research. However, Elliott & Shanahan noted that smokers were more familiar with the text only labels and that new versions of text-only warnings were seen as “more of the same” (p. 30). Similarly, they made a number of comments in the latter two reports about the relative difficulty respondents had in “deflecting” the graphics relative to the text-only warnings. Both conclusions would appear to support an argument that the text-only warnings were less impactful than the graphics, and wearout or increasing familiarity, could explain this finding. On the basis of the group discussions undertaken, Elliott & Shanahan concluded that the health warning labels had become less noticeable over time, despite the fact that awareness of them remained high. More specifically, they reported that respondents thought the warning images had lost their initial novelty and attraction and had become “old hat” (p. 18). Expert stakeholders concurred with this view, but it is incorrect to state that this view was only apparent among expert stakeholders.

Overall, Elliott & Shanahan’s work provides a comprehensive background to health warnings and the way in which these are used and responded to by smokers. However, the studies did not examine specific behavioural responses from smokers, although the variables offer insights into the factors smokers stated were likely to affect their behaviour. In our opinion, many of the criticisms levelled at Elliott & Shanahan’s work by Todd cannot be tested without direct recourse to the interview transcripts, and even these may not allow her concerns to be addressed. In the absence of this information, we believe it is appropriate to consider the researchers’ credentials. We note that the key personnel are both registered psychologists and qualified practising market researchers. Given the backgrounds of the researchers, it seems likely they were cognisant of the issues Todd raised and that they had the practical experience and training to address these. Todd has provided no specific evidence of the factors she alleges *may* have affected the research; many factors *may* affect research but users of research need to deal with actual evidence of these factors not hypothetical assertions about their existence.

Dixon’s main arguments were that there is no evidence the text-only warnings have become less effective, an argument he based on the similar levels of awareness detected in the 1996 and 2000 surveys. However, as we have noted, the survey data allow for alternative explanations more consistent with Elliott & Shanahan’s conclusions. Moreover, measures of awareness are not the same as measures of wearout, and the fact that awareness has remained at similar levels is not evidence that a message or format has remained impactful. Instead, awareness may be a consequence of learning and it should not be interpreted to mean that a message or format has retained its salience.

In summary, while Elliott & Shanahan’s work did not establish that warnings featuring graphics would decrease the proportion of smokers or reduce initiation rates among non-smokers, nor did it establish

that maintaining the status quo was the most effective health policy option. Given smokers' increased agreement with the need for stronger warnings and the evidence from focus groups that smokers see graphic images as more likely to cue quitting related behaviours, Elliott & Shanahan's work appears to support the introduction of pictorial warning labels.

New Zealand Research

Barwick et al, 1995

Although Barwick *et al's* report was published a decade ago, it contains a detailed and dispassionate review of a wide body of relevant literature. The report comprises three sections, of which one deals specifically with health warnings and information. Barwick *et al* suggested four reasons are generally given to support the introduction of health warnings: consumers' right to access information that enables them to make more informed choices; government's need to demonstrate their opposition to smoking; the role played by health warnings in supporting overall smoking education programmes, and, most controversially, their ability to influence smoking behaviour.

While Barwick *et al* were not able to identify a clear relationship between health warnings and behaviour, the research they reviewed indicated that cigarette packages were the first source of health information for smokers, and the third most important information source overall. This suggests that visually impactful on-pack warnings may increase the salience and personal relevance of the health risks associated with smoking, although this does not necessarily mean they will also affect smokers' behaviour.

The research reviewed by Barwick *et al* was conducted prior to the introduction of pictorial health warnings by any jurisdictions and their conclusions about the extent to which information provided in health warnings actually warns smokers relate only to text warnings. Thus findings that suggest text-based on-pack warnings have little impact cannot necessarily be generalised to pictorial warnings and may simply indicate that the text-based warnings examined suffered from wearout and needed refreshing with new warnings.

Although Barwick *et al* concluded that the research they reviewed had not identified a clear empirical relationship between health warnings and smoking behaviours, they did not believe this criterion had to be met to justify the use of health warnings. In their view, consumers' expectations that they would receive information about products, and governments' role in providing information to citizens, justified the use of health warnings.

Because Barwick *et al* were not able to assess research relating to pictorial health warnings, their findings cannot answer the question of whether these should replace the current text-based warnings, which the tobacco industry argues are sufficient. We believe cross-sectional research designed to evaluate alternative warning label formats (discussed in more detail in section 2.2.4) would be an appropriate means of addressing this question.

Waa et al (2004)

The research reported in Waa *et al* was part of a wider study designed to examine second hand smoking. The survey involved telephone interviews with 1507 New Zealanders drawn from the general population and a further 902 interviews with Maori (an overall response rate of 61% was achieved). One section of the questionnaire examined health warnings and respondents were asked about their awareness of on-pack health warnings; 97% of daily smokers were aware of these. Daily smokers' recall of specific health warnings varied: 63% recalled "smoking when pregnant harms your baby" while 11% recalled "smoking is addictive". Interestingly, the health warning that elicited the highest levels of agreement from both the general population sample (GPS) and the Maori population sample (MPS) was "smoking is addictive". Levels of agreement with all messages ranged from 81% to 95% among both samples.

When asked the effect health warnings had on their behaviour 75% (79% MPS) indicated they had thought about quitting; 67% (73% MPS) had made a quit attempt; 57% (62%) had reduced their cigarette consumption, and 14% (22% MPS) had phoned the Quitline number. Awareness of the Quitline number was generally low; 31% of daily smokers were aware of the purpose of the number while nearly 50% stated they did not know what this was.

Todd criticised this report for drawing conclusions she claimed were not adequately supported by the evidence presented. In particular, she criticised the question wording used to explore respondents' behavioural response to the health warnings and argued that this "forced [everyone] to respond that they had been influenced in some way". This argument is not correct. We assume Todd is referring to Q39 of the study, which asked respondents: "Have you ever done any of the following **because** of the health warnings on cigarette packages?" Respondents were then read a list of behaviours to which they could respond yes/no/don't know. In our opinion, the structure of the question does not force respondents to indicate they had performed any of the behaviours listed.

Todd also argued that social desirability error could have affected the responses provided, although she offers no evidence of specific estimates that she believes were affected. Again, the responses to the behaviours listed varied, which suggests respondents felt able to indicate freely which of these they had performed. Similarly, the fact that only 14% (22% MPS) stated they had called the Quitline is not inconsistent with the finding that 30% (28% MPS) were aware of the purpose of the number.

As Todd noted, unprompted recall of the messages was higher among lower age groups, although agreement with these was lower. This finding could support arguments that young people do not find the current text-only messages salient and is consistent with the suggestion that introducing graphics could increase awareness of health issues and agreement with statements positing a link between smoking and adverse health effects. Interestingly, although agreement with the warning "smoking is addictive" was high relative to agreement with other health warnings, awareness of this

warning was lower. One explanation of this is that the health warnings are not effectively communicating with smokers.

Dixon also commented on Waa *et al*'s research and argued that the research findings indicate a high awareness of the health warnings, which prompted larger proportions of smokers to consider quitting or attempting to quit. As a consequence, he concluded that the current health warnings appear to be very effective in influencing smokers' attitudes. However, closer analysis of Waa *et al*'s findings suggests that although claimed awareness was high, actual recall of warnings was considerably lower. Although the data in Table 7 suggest smokers have responded to the health warnings, a number of the variables tested (such as thinking about quitting) cannot be linked directly to behaviour. Those variables that do have a closer link to behaviour show varying patterns – while 67% (73% MPS) claim to have attempted to quit and 57% (62% MPS) state they have reduced the amount of tobacco they consume, only 14% (22% MPS) noted they had actually called the Quitline.

BRC Research (May and August 2004)

BRC conducted two separate studies into text and pictorial warnings. The first of these involved 100 face-to-face interviews (50 smokers, 17 recent quitters and 27 non-smokers), designed to explore which of a range of health warnings were most and least likely to impact on smokers and communicate the effects of smoking. In addition, the research investigated design aspects of the messages; the extent to which the health warnings conveyed believable information, influenced beliefs and attitudes towards smoking, provided an avenue for smokers to seek quitting help, and affected behaviour.

The research classified warnings into five themes: children, second hand smoke and role-modelling; physical health; poison; quitting help, and addiction, and tested a range of messages that related to these themes. The methodology used a range of techniques to address the objectives and the questionnaire was pre-tested prior to the full interviewing round.

The BRC report identified the messages most likely to have an impact, or best communicate the themes examined. When asked to explain their selection, respondents often commented on the image used, the emotional connection these generated, and the clear demonstration of harm they provided. While respondents offered suggestions they felt could improve some images, the BRC report suggested these nevertheless had a very strong impact on them and concluded that "graphic images with shock-factor" (p. 28) were seen as effective, as were the use of colour and bold text.

Respondents were also asked to indicate their level of agreement with statements about the likely effect of the images they were shown on smokers (or sub-groups of smokers). Although the sub-samples on which the estimates are based are small, the survey findings suggest the images would

be likely to encourage smokers to reduce the amount of tobacco they consumed or to think about or attempt to quit.

The second BRC study involved 8 mini group discussions convened to conduct more specific tests of design elements. In particular, the research examined the balance between pictorial and text warnings; whether pictorial warnings were considered more impactful than text warnings; the effect of the size of the health warnings, and the use of colour in the health warnings. BRC found that packs featuring a large image (50% of the front display area) were rated most impactful by respondents, who found the larger image more prominent and better able to stand out from the brand imagery that featured on packs. As a result, they concluded that pictorial warnings had more impact than text only warnings, and the larger the graphic image, the greater the likely impact. They also drew a number of conclusions about the position of the image, the impact of different colour combinations, the use of side messages, and other attributes, such as simplicity, brevity, and shock, which were likely to increase message impact.

Both Todd and Dixon levelled several criticisms at the BRC research, which Todd suggested assumed rather than tested the effect of health warnings on behaviour. She argued that until the relationship between health warnings and behaviour had been established, research designed to test executional factors was premature.

Todd also criticised the small samples involved and the use of face-to-face interviews. However, given that the research involved extensive use of visual stimuli, including mock-up packs, we believe face-to-face interviews were an appropriate survey mode. Todd raised further concerns about interviewer and social desirability error; for reasons that she did not explain in detail, she questioned the use of senior BRC researchers to undertake the research. In our opinion, use of senior researchers is entirely appropriate and would be likely to decrease, rather than increase, any potential for interviewer error. Given that the interviewers were all experienced, the potential for social desirability error to have affected respondents would also, in our opinion, be reduced. Again, Todd offered no examples of where she believed social desirability error had affected the data and, without specific instances to evaluate, it is difficult to see the basis of her concerns.

Todd also appeared to suggest that, because respondents saw television as a more "effective" communication channel than cigarette packages, the latter could not effectively communicate information to smokers. Respondents' ranking of television as more effective is unsurprising, given that they would have been conditioned to receiving information via television for several years. However, as Allen & Clarke's review noted, cigarette packages have a unique presence at the point at which tobacco consumption occurs, thus it is logical to assume information on cigarette packs reaches

respondents. The high levels of awareness that cigarette packages contain health warnings corroborate this interpretation.

Dixon argued that the BRC research was similar in style to work undertaken by Health Canada and by Elliott & Shanahan in Australia and suggested that its emphasis on communication variables offered no insights into smokers' behavioural response. Dixon also commented on an assertion made in the BRC report that appears to pre-empt the criticism that the research measures are not explicitly behavioural. BRC stated that the rationale for health warnings is wider than behaviour change, although, as Dixon noted, the report did not expand on this view. He thus interpreted their comment to imply that if health warnings did not aim to change behaviour, their purpose must be to improve awareness. Because the research evidence suggests smokers are already aware that cigarette packages contain health warnings, Dixon argued that justification of pictorial health warnings because they improve awareness is flawed. However, Waa *et al's* work showed that although smokers know health warnings feature on cigarette packages, fewer could recall specific examples of these. Given this discrepancy between respondents' general awareness of health warnings and their ability to recall these, Dixon's argument that awareness has already been achieved is debateable.

In addition, Dixon did not address the question of whether alternative approaches might be more effective at increasing awareness, particularly specific awareness of health risks. The Canadian research evidence revealed that since the introduction of pictorial warning labels, the proportion of smokers stating they have seen information about the health effects of smoking has increased sharply. Top of mind and aided recall of several specific health effects also increased following the introduction of pictorial warning labels. It is not possible to comment on the behavioural outcomes associated with the Canadian regulations, since Dixon did not report specific results from the later data waves that he accessed. If this information were to be placed in the public domain, an assessment of self-reported behavioural consequences over a longer time period would be possible.

In summary, Dixon and Todd did not support their criticisms of the design and implementation of the BRC research with specific examples or logical evidence. Nevertheless, the BRC research does not specifically examine the effect the health warnings would have on smokers' actual behaviour. While the pictorial warnings were clearly more impactful, this is not in itself evidence that they would reduce smoking initiation or tobacco consumption.

3. Conclusions

3.1 Criticisms of the MoH Consultation Document

The tobacco industry submissions on the Ministry of Health consultation document raised several questions about the quality of the research used to inform the MoH's thinking. In particular, they argued that the research findings to date do not constitute a compelling case for introducing pictorial health warnings. To support this argument, they alleged that:

- The studies relied on by the MoH are methodologically flawed and unreliable;
- The outcome variables are typically awareness, knowledge or attitude, none of which offer robust insights into people's likely behaviour,
- Only longitudinal research with behavioural outcomes would allow an estimation of the effect of pictorial health warnings on smokers' and non-smokers' behaviour.

We address these alleged deficiencies below.

Methodological Flaws

Todd suggested that social desirability error could have affected estimates in a number of the studies reviewed, although she did not provide any specific instances to illustrate her concerns. As we noted earlier, social desirability error may affect surveys that investigate sensitive or topical subjects, although it does not inevitably affect all surveys dealing with these issues.

One means of testing whether social desirability error has affected survey estimates involves comparing these to a known population parameter. However, where the variables tested are primarily cognitive and affective, this is not possible, and there are no external measures against which we can assess the survey estimates.

We have examined the questionnaires included in research reports (we note these were not available for all the studies reviewed) and, where we can make an assessment, we believe the questions were worded neutrally to allow a full range of responses to be given. However, we note that even well-designed questions may elicit responses affected by social desirability error. Thus, while we have seen no evidence that this error affected the surveys reviewed or had a material effect on the results reported, we cannot completely eliminate the possibility that it did affect respondents' answers.

Todd also commented on the low response rates achieved by some of the surveys reviewed and, where these were reported, it is clear that the response rates varied considerably (and the methods used to calculate these may also have varied). Low response rates increase the potential for non-

response error to affect the estimates. Because there are rarely known population parameters against which to compare the estimates, it is typically not possible to assess the presence or extent of any non-response error. As we noted in our review of the papers, the surveys that achieved low response rates often used sophisticated sampling procedures to ensure the wider population was matched on known demographic characteristics. While there is no apparent reason why individuals in a particular demographic who did not respond to the survey should differ from those in the same demographic who did respond, we cannot rule out the possibility that differences between respondents and non-respondents did exist.

Todd criticised the demographic skew she alleged was present in the research studies reviewed and the scarcity of information from respondents aged less than 18 years. We note that many of the research studies did include data from younger age groups (Envionics, 2001a; Kaiserman *et al*, 2004; Shanahan, 2000; Elliott & Shanahan, 2002; 2003; Waa *et al*, 2004).

Overall, from the information we had, we found no reason to believe that the research studies were not well-designed and conducted. As far as we could assess, the researchers were experienced and well-qualified, and the surveys appear to have been competently designed, implemented and reported.

Outcome Variables

Both Dixon and Todd criticised the MoH for unclear definitions of "effectiveness" and for what they believed was an excessive reliance on cognitive rather than behavioural outcome measures. We believe the definitions of "effectiveness" did vary from one study to another, although we note that as many studies were not designed to be part of the same research programme, this variation is not unexpected. Thus, while some studies sought to examine the communication effectiveness of health warning messages, others examined their effect on respondents' recall, knowledge and self-reported behaviour. These differences were properly reflected in the varying measures of effectiveness set out in the research objectives for each study.

Many of the studies we reviewed examined respondents' awareness of health warnings, the extent to which they agreed or disagreed with the information provided in these, and the effect they believed these warnings had on their own behaviour. Use of cognitive variables, such as attitudes, cannot be assumed to be equivalent to actual behaviour. Although self-reported behaviours are arguably closer proxies of actual behaviour, even these should not be assumed to be measures of actual behaviour.

Ideally, research into the effect of pictorial warning labels would provide clear evidence that these labels reduced the proportion of smokers and the rate at which they consumed tobacco, and would demonstrate a reduction in the number of new smokers. However, because a range of variables may

affect behaviour, identifying direct causal relationships may be difficult, if not impossible. As an alternative, Dixon and Todd have recommended longitudinal research, which we discuss in more detail below.

Longitudinal Research

As we have noted, it is not possible to estimate the actual effect of pictorial warning labels on smokers' (or non-smokers') behaviour until the warnings have actually been introduced. The research undertaken by Gospodinov and Irving (2003; 2004) used data collected in the Canadian Tobacco Use Monitoring Survey (which itself relied upon self-reported behaviour rather than actual behaviour) to estimate the effect pictorial warning labels had on Canadian smokers' behaviour. Dixon argued that the consultation document placed insufficient emphasis on the findings from this research, which involves a longitudinal study. However, the limitations affecting other studies relying on self-reported data would appear to apply equally to Gospodinov and Irving's research.

Unlike the other studies discussed, which concluded that pictorial warnings were associated with changes in smokers' cognitive and quasi-behavioural responses, Gospodinov and Irving reported no evidence of a decline in prevalence although they found "surprising" evidence of a decline in consumption. In fact, both measures they examined (prevalence and consumption) decreased, even if the size of the decrease was significant for only one of the variables (consumption). In our opinion, these findings are not inconsistent with the conclusion that pictorial health warnings are associated with a change in smokers' self-reported behaviours.

We believe that longitudinal behavioural research, of the type alluded to in the tobacco industry's submissions, would not necessarily enable indisputable conclusions to be drawn about the impact of pictorial health warnings on a range of smoking-related behaviours. We note that Todd and Dixon criticised the studies involving longitudinal research (Borland & Hill, 1997; Hammond *et al*, 2003; 2004). Longitudinal research would be difficult, time-consuming and expensive to undertake, and the results would be likely to provoke rather than resolve dispute over the effects of pictorial health warnings. Furthermore, as we noted earlier, longitudinal research into the behavioural effects of an intervention can only occur once the intervention has been introduced, which would seem to obviate the need for any research to test whether it should be implemented. Even assuming longitudinal research could be undertaken, dispute over the relationship between any observed decline in smoking prevalence or consumption rates and specific interventions would be likely to occur.

Alternative Research Approach

To address this problem, we suggest that well-designed cross-sectional research would enable a comparison of proposed measures and existing measures. The work undertaken by Liefeld could be replicated, although, if this suggestion is adopted, we recommend changing the dependent variable

used from “encouragement”, assessed using an ordinal scale, to likelihood or probability, and assessed using a metric-level scale such as the Juster Scale. We have adopted a similar approach in pilot work (attached as Appendix A) and we believe this work could also provide the basis for more extensive research into the effects of pictorial warning messages.

However, we note that the research we have outlined would involve comparing respondents’ estimated likelihood of performing a range of cessation related behaviours (reducing the number of cigarettes smoked, phoning the Quitline). When provided with a defined context, we have found respondents can provide very accurate estimates of their likely behaviour, although we note that this measure is still one step removed from actual behaviour. Nevertheless, given the difficulty of obtaining behavioural data, we believe this approach would provide a sound basis on which to make policy decisions.

Evaluation of Pictorial Warning Labels

We have also been asked to consider how the effects of pictorial warning labels could be evaluated, should they be introduced. Ideally, evaluation research would be able to assess the specific contribution an intervention made. In practice, any regulatory change is likely to be accompanied by media discussion, advertising campaigns, and other cessation strategies designed to maximise the potential effects the change may bring. As a result, defining the precise contribution made by pictorial warning labels is likely to be very difficult, if not impossible.

Evaluation research should also test the objectives set for the intervention. If these relate to communication goals, then Health Canada’s approach could be adapted for use in New Zealand. The research reported by Kaiserman *et al* involved several waves of data collection and explored a range of awareness, knowledge and cognitive variables, as well as examining respondents’ self-reported behaviour. On-going tracking would provide information on the communication effectiveness of the new warning labels, and how smokers describe themselves as having responded to these.

If the objective was to increase cessation-related behaviours, the research could include a benchmark study that estimated smokers’ likelihood of quitting once the labels had been introduced and a follow-up study where the same sample and a control sample (to assess any testing effects) were interviewed and behaviour changes monitored. However, as noted, while these results could test whether the introduction of pictorial warning labels was correlated with an increase in cessation behaviours, it could not test whether the labels had caused the effects observed.

3.2 Summary Comments

The studies reviewed are diverse in their design and implementation, although some common themes are evident in them. The Canadian research, much of which was conducted after the introduction of pictorial warning labels, suggested the introduction of graphics has made the labels more impactful. After the change to the new warning labels smokers reported increased awareness of specific health risks and higher agreement with the warnings. The proportion who reported they were likely (or more encouraged) to engage in cessation related behaviours also increased. Econometric analyses also supported the conclusion that the prevalence of smoking and the number of cigarettes consumed had decreased, although the former finding was not statistically significant.

The Australian research suggested smokers supported the introduction of stronger warning labels on cigarette packages. The research found some evidence that the text warnings introduced in 1996 were associated with an increase in the number of smokers reporting they were concerned about the health consequences of smoking. However, there was no evidence that respondents felt the new text warnings had led them to smoke less or quit. The proportion of smokers and non-smokers who supported stronger health warnings increased from 1996 to 2000.

Subsequent qualitative work suggested the impact of the text-only warnings had declined and the graphics tested with respondents were consistently seen as more impactful, although respondents' reactions to the proposed warning messages varied across demographic groups. Nevertheless, although some groups had entrenched attitudes and behaviours, all felt the pictorial warnings would be difficult to overlook, and that they would act as a deterrent to new smokers and those contemplating smoking.

The New Zealand research suggested that unprompted awareness of health warning messages is variable, although agreement with their content was high. Research designed to compare text and pictorial warning messages found the latter evoked more emotional responses. The research also concluded that pictorial warnings would be likely to prompt smokers to reduce their tobacco consumption and increase the proportion who attempted to quit.

Although each of the studies discussed has limitations, we believe they were generally well designed and conducted and, when viewed holistically, they suggest that pictorial warning labels will be more impactful and credible than text-only warning messages. For health researchers, these findings provide compelling evidence that pictorial warning labels should be mandatory, since they have improved the communication effectiveness of text-only warning messages. The research evidence also suggests pictorial warning labels are associated with an increase in the proportion of smokers who report cessation-related behaviours. However, for the tobacco industry, the question is not so much whether the pictorial warnings represent an improvement, but whether the status quo is

currently adequate. The research we have reviewed cannot answer these questions, which are ultimately philosophical and political.

In summary, we have found that:

- The flaws alleged to have affected the studies reported in the MoH consultation document are typically often presented in general terms without detailed evidence of their effects. As far as we can tell, the surveys discussed in the MoH consultation document were generally well designed and conducted. We have seen no evidence that the estimates reported have been materially affected by the errors outlined in the tobacco industry submissions, although it is not possible to assess all the alleged errors.
- Much of the research cited used changes in awareness or other cognitive variables as outcome measures. These measures cannot be assumed to lead to behaviour, nor do they necessarily offer insights into respondents' future behaviour.
- Arguments that evidence of behaviour change is necessary to support proposals to introduce pictorial warning labels are logically flawed as they pre-suppose the introduction of the measures they propose to assess.
- Although behaviour change is a desired outcome of proposals to introduce pictorial warning labels, we believe it would be extremely difficult to undertake research that could isolate the specific contribution of pictorial warning labels to smoking cessation or non-initiation, and that would be accepted by all parties.
- In the absence of behavioural measures, it is nevertheless possible to compare alternative warning labels and to obtain estimates of their likely effects on respondents' behaviour.
- If pictorial warning labels were introduced, appropriate evaluation measures would depend on the objectives they were intended to realise. Communication objectives could be assessed using a similar methodology to that employed by Health Canada. Behavioural objectives could be assessed using a phased study that incorporated control sub-samples.

Bibliography

- Allen & Clarke Policy and Regulatory Specialists Ltd. (2003). *Options for the Regulation of Tobacco Products*. Unpublished report to the New Zealand Ministry of Health, July.
www.moh.govt.nz or www.allenandclarke.com
- Barwick, H., Bergham, P., Burns, J., et al. (1995). Smoke-free Issues. An analysis of key issues underpinning proposed amendments to the smoke-free environments legislation. Unpublished report, available from the New Zealand Ministry of Health.
- Boddewyn, J. J. (1989). There is no convincing evidence for a relationship between cigarette advertising and consumption. *Br J Addict*, 84, 1255-61.
- Borland, R. and Hill, D. (1997). Initial impact of the new Australian tobacco health warnings on knowledge and beliefs. *Tobacco Control*, 6 (4), 317-25.
- BRC. (2004). Smoking Health Warnings Study: The effectiveness of different (pictorial) health warnings in helping people consider their smoking-related behaviour. Wellington: BRC Marketing & Social Research. URL:
<http://www.ndp.govt.nz/publications/healthwarnings.html>
- Chetwynd, J., Coope, P., Brodie, R. J., and Wells, E. (1988). Impact of cigarette advertising on aggregate demand for cigarettes in New Zealand. *British Journal of Addiction*, 83, 409-14.
- Elliott and Shanahan Research (1996). Evaluation of health warning labels on tobacco products. Unpublished report prepared for the Commonwealth Department of Health and Family Services.
- Elliott and Shanahan Research (2000). Evaluation of the health warnings and health messages on tobacco products. Report for the Australian Dept of Health and Ageing
(www.health.gov.au/pubhlth/strateg/drugs/tobacco/warnings.html)
- Elliott and Shanahan Research (2002). *Developmental Research for New Australian Health Warnings on Tobacco Products*, Stage 1, prepared for Population Health Division, Department of Health and Ageing, Canberra.
- Elliott and Shanahan Research (2003). *Developmental Research for New Australian Health Warnings on Tobacco Products Stage 2*. Canberra: Commonwealth of Australia.
http://www.health.gov.au/pubhlth/strateg/drugs/tobacco/warnings_stage2.pdf.
- Enviroics (1999). Focus Group Report to Health Canada - Office for Tobacco Control on Attitudes Towards 'Light' and 'Mild' Cigarettes. PN4361.
- Enviroics (2001a). Baseline Surveys: The health effects of tobacco and health warning messages on cigarette package (final report for Health Canada).
- Enviroics (2001b). Evaluation of New Warnings on Cigarette Packages. Prepared for Canadian Cancer Society.
http://www.cancer.ca/vgn/images/portal/cit_776/35/20/41720738niw_labelstudy.pdf.
- Gospodinov N., and Irvine I. (2003.) Global Health Warnings on Tobacco Packaging: Evidence from the Canadian experiment. Unpublished study.
- Gospodinov, N., and Irvine, I. (2004). Global Health Warnings on Tobacco Packaging: Evidence from the Canadian Experiment, July 2004 version supplied by I Irvine to Health Canada
- Hammond, D., Fong, G., McDonald, P., et al. (2003) Impact of the graphic Canadian warning labels on adult smoking. *Tobacco Control*, 12, 391-395.
- Hammond, D., Fong, G., McDonald, P., et al. (2004). Graphic Canadian cigarette warning labels and adverse outcomes: evidence from Canadian smokers. *Am J Public Health*, 94(8): 1442-5.
- Kaiserman, M., Malomaski, Illing, E., et al. (2003) The evaluation of Canada's health warning messages: 18 month follow-up. Presented at the 12th World Conference on Tobacco or Health, Helsinki.
- Laugesen, M., and Meads, C. (1991) Tobacco advertising restrictions, price, income and tobacco consumption in OECD countries, 1960-1986. *British Journal of Addiction*, 86, 1343- 1354.
- Liefeld, J. P. (1999) The relative importance of the size, content and pictures on cigarette package warning messages. Report for Health Canada (www.hcsc.gc.ca/hecs-sesc/tobacco/research/archive/index.html)
- Ministry of Health. 2004. Review of the Smoke-free Environments Regulations 1999: Labelling of tobacco products, tobacco product content regulation, disclosure by tobacco companies, regulation of product descriptors: Consultation document. Wellington: Ministry of Health.

- Nilsson, T. (1999). Legibility and visual effectiveness of some proposed and current health warnings on cigarette packages. Report for Health Canada. (www.hc-sc.gc.ca/hecs-sesc/tobacco/research/archive/index.html)
- Stewart, M. J., (1992). Tobacco Consumption and Advertising Restrictions: A Critique of Laugesen and Meads (1991). *International Journal of Advertising*. 11, 97-118.
- Waa, A., Gillespie, J., and Afzal, R., (2004) Influence on Smoking Behaviour and Perceptions of Cigarette Packaging in New Zealand: Preliminary findings. Unpublished report to the New Zealand Ministry of Health.
<http://www.ndp.govt.nz/publications/packagingandlabelling.html>.