IARC HANDBOOKS OF CANCER PREVENTION

Tobacco Control

IARC

International Agency for Research on Cancer
World Health Organization

Volume 12

Methods for Evaluating Tobacco Control Policies

IARC
2008

This Handbook was made possible thanks to the generous funding by the Ministère de la Santé, de la Jeunesse et des Sports

5.5 Measures to evaluate the effectiveness of tobacco product labelling policies

Background

The cigarette package serves as the cornerstone of tobacco marketing and advertising campaigns (Slade, 1997; Pollay, 2001). Package design helps to reinforce brand imagery communicated through other media and plays a central role in retail marketing. The importance of cigarette packaging only increases as other forms of marketing are restricted, as indicated in the following quote from a Philip Morris executive: "Our final communication vehicle with our smoker is the pack itself. In the absence of any other marketing messages, our packaging...is the sole communicator of our brand essence. Put another way—when you don't have anything else—our packaging is our marketing." (Alechnowicz & Chapman, 2004).

Governments in many jurisdictions have begun to apply greater restrictions on tobacco labelling. As much as half of the package is now used by regulators to communicate the health effects of smoking. Governments have also begun to prohibit packaging elements that are deemed to be misleading to smokers. As a consequence, labelling policies have begun to alter the traditional appearance of the cigarette package.

The importance of tobacco labelling policies is highlighted in Article 11 of the WHO FCTC (WHO, 2003). Article 11 sets international standards for packaging and labelling of tobacco products in three broad categories: 1) mandatory health warnings; 2) restrictions on brand descriptors, such as the use of "light" and "mild"; and 3) information on cigarette contents and emissions[1] (Figure 5.26).

Health warning labelling

Cigarette packages in the vast majority of countries carry a health warning (Aftab *et al.,* 1999). However, the position, size, and general strength of these warnings vary considerably across jurisdictions. FCTC Article 11 requires that package health warnings must cover at least 30% of the package surface and be "large, clear, visible, and legible" (WHO, 2003). Beyond these minimum requirements, Article 11 also states that warnings "should" cover 50% or more of a package's principle surfaces, and "may" be in the form of pictures.

To date, at least eight countries have implemented picture-based health warnings that meet the FCTC's "recommended" standard (see Figure 5.27). A number of other jurisdictions, including the European Union, have recently implemented prominent text warnings which meet the minimum FCTC standard. More obscure text warnings remain in many other markets, including the USA, China, and Russia.

Constituents & emissions labelling

There is general agreement that tobacco packaging should include at least minimal information about some of the hazardous and addictive constituents in tobacco and tobacco smoke. FCTC Article 11 requires that packages contain "information on relevant constituents and emissions of tobacco

[1]Tobacco labelling policies apply to a broad range of tobacco products, including a range of combustible products, such as cigars, and the packaging of loose or "fine cut" tobacco, as well as non-combustible tobacco products. However, much of this section will focus on labelling policies for factory-made, "pre-packaged" cigarettes given that they are the primary target of labelling policies, and the area in which most research has been conducted. Labelling policies for other types of products will be described briefly in a separate section to follow.

1. Each Party shall, within a period of three years after entry into force of this Convention for that Party, adopt and implement, in accordance with its national law, effective measures to ensure that:

(a) Tobacco product packaging and labelling do not promote a tobacco product by any means that are false, misleading, deceptive or likely to create an erroneous impression about its characteristics, health effects, hazards or emissions, including any term, descriptor, trademark, figurative or any other sign that directly or indirectly creates the false impression that a particular tobacco product is less harmful than other tobacco products. These may include terms such as “low tar”, “light”, “ultra-light”, or “mild”; and

(b) Each unit packet and package of tobacco products and any outside packaging and labelling of such products also carry health warnings describing the harmful effects of tobacco use, and may include other appropriate messages. These warnings and messages:

(i) shall be approved by the competent national authority,
(ii) shall be rotating,
(iii) shall be large, clear, visible and legible,
(iv) should be 50% or more of the principal display areas but shall be no less than 30% of the principal display areas,
(v) may be in the form of or include pictures or pictograms.

2. Each unit packet and package of tobacco products and any outside packaging and labelling of such products shall, in addition to the warnings specified in paragraph 1(b) of this Article, contain information on relevant constituents and emissions of tobacco products as defined by national authorities.

3. Each Party shall require that the warnings and other textual information specified in paragraphs 1(b) and paragraph 2 of this Article will appear on each unit packet and package of tobacco products and any outside packaging and labelling of such products in its principal language or languages.

4. For the purposes of this Article, the term “outside packaging and labelling” in relation to tobacco products applies to any packaging and labelling used in the retail sale of the product.

WHO (2003)

Figure 5.26 WHO FCTC Article 11: *Packaging and labelling of tobacco products*

products as defined by national authorities.” At present, however, national authorities have taken much different approaches to labelling constituents and emissions, and there remains considerable disagreement regarding what should be considered “relevant” information.

The current regulatory practice in many jurisdictions is to require manufacturers to print levels for three emissions in the mainstream smoke: tar, nicotine, and carbon monoxide (CO). Emission levels are generated by machine-smoking cigarettes according to a standard set of puffing conditions; typically the International Standards Organization (ISO) method, which serves as the current international standard. However, in light of research indicating that the tar and nicotine levels generated under the ISO testing method are unrelated to individual levels of exposure or risk (Burns *et al.*, 2001), there are growing calls from within the tobacco control community for the ISO numbers to be removed from packages. Some jurisdictions have supplemented the ISO numbers with additional emission information. For example, Canada increased the list of emissions that must be reported

Figure 5.27 Picture-based warnings

(by adding benzene, formaldehyde, and hydrogen cyanide to tar, nicotine, and CO), and requires a second set of numbers from a more intensive machine smoking method for each emission (Figure 5.28). Other jurisdictions have replaced quantitative emission values with descriptive, non-numerical information on hazardous emissions and toxicants. A consensus has yet to emerge on "best practices" for this area of tobacco labelling policy.

Brand descriptor labelling

Tobacco manufacturers incorporate a variety of common terms into the names of their cigarette brands. Words such as "light" and "mild" are ostensibly used to denote flavour and taste; however, "light" and "mild" brands are often promoted as "healthier" products and are typically applied to brands that generate lower machine levels of tar (Pollay, 2001; Pollay & Dewhirst, 2002). Not surprisingly, "light" and "mild" brands are perceived by many consumers to deliver less tar and lower risk than "regular" or "full flavour" varieties despite evidence to the contrary (Ashley *et al.,* 2001; Shiffman *et al.*, 2001).

A growing number of jurisdictions, including Brazil and the European Union, have prohibited the use of "light" and "mild" on packages. Similar prohibitions are proposed in FCTC Article 11: "tobacco product packaging and labelling do not promote a tobacco product by any means that are false, misleading, deceptive or likely to create an erroneous impression about its characteristics, health effects, hazards or emissions, including any term, descriptor, trademark, figurative or any other sign that directly or indirectly creates the false impression that a particular tobacco product is less harmful than other tobacco products. These may include terms such as "low tar," "light," "ultra-light," or "mild". Although there is evidence to suggest that other packaging elements, such as the use of colour, may also create misleading perceptions of risk (Wakefield *et al.*, 2002), "light" and "mild" descriptors are the only packaging elements to be restricted to date.

Methodological issues in evaluating tobacco labelling policies

Evaluating tobacco labelling policies presents several unique challenges; this section reviews some of the principal methodological and analytical considerations.

"Alternative" tobacco products:

Labelling policies have generally been designed with factory-made, pre-packaged cigarettes in mind. However, a substantial proportion of tobacco users throughout the world use tobacco products that are either packaged in a different way, or have no manufactured packaging at all. This has important implications for patterns of exposure to health warnings. For example, consumers who buy loose or fine-cut tobacco, without any manufactured packaging, may not be exposed to product health warnings. Even consumers who buy fine-cut tobacco, sold in government-mandated packaging, will have different patterns of exposure than those who smoke manufactured cigarettes, and who are likely to be exposed to the warnings each time they reach for the package. As a result, studies conducted in markets with a considerable proportion of fine-cut tobacco sales, such as the United Kingdom, New Zealand, and Thailand, may need to stratify for fine-cut versus manufactured or mixed use. Smuggled or contraband cigarettes may also alter patterns of exposure in cases when the contraband product is not manufactured to the same labelling specifications.

Issues in attribution: dealing with multiple sources of health information:

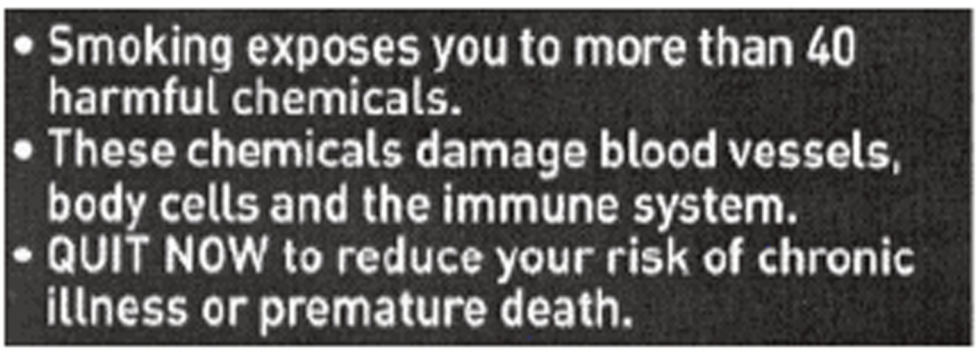
Health behaviours with multiple determinants present a challenge to policy evaluation. The problem of attribution is particularly acute for health warning labels. First, labelling policies are often implemented simultaneously with other tobacco control measures, including increases in taxation and smoke-free policies. As a result, it is difficult to isolate the effect of an individual policy on overall prevalence. Second, many of the specific themes and messages in labelling policies are communicated through other sources.

European Union (United Kingdom): Three ISO emissions

"Toxic emissions / unit:" "Tar" 14 - 35 mg,
Nicotine 1.1 - 2.7 mg, Carbon monoxide 14 - 30 mg,
Formaldehyde 0.055 - 0.14 mg,
Hydrogen cyanide 0.14 - 0.36 mg, Benzene 0.043 - 0.097 mg

"Émissions toxiques / unité:" "Goudron" 14 - 35 mg,
Nicotine 1,1 - 2,7 mg, Monoxyde de carbone 14 - 30 mg,
Formaldéhyde 0,055 - 0,14 mg,
Acide cyanhydrique 0,14 - 0,36 mg, Benzène 0,043 - 0,097 mg

Canada: Six ISO emissions and 'Health Candada Intense' emission



Australia: Descriptive information

Figure 5.28 Constituent labelling policies in the European Union, Canada and Australia

Mass media campaigns and health professionals often target the same health effects, particularly with regards to common diseases such as cancer and cardiovascular disease. The impact of package-based labelling policies may also be confounded with health warnings in other settings. Various jurisdictions require health warnings in retail outlets and warnings on print advertisements for tobacco products. Third, perceptions of risk and health knowledge are influenced by an inter-related set of factors at the individual, social, and environmental level. Few studies are able to measure more than a small number of these factors within a single study and none can fully isolate the contributions of each. These realities underscore the importance of the methodological features described in Section 2.1. In addition, environmental scans of other mass media campaigns and policy interventions can provide important context.

"Wear-out" and impact over time:

It is widely accepted that the salience of advertising and health communications is typically greatest upon first exposure (Bornstein, 1989; Henderson, 2000). The initial impact of comprehensive labelling policies, such as the introduction of large graphic warnings on packages, is often magnified by media coverage. As a result, measures of effectiveness are likely to be strongly associated with the implementation date. This has implications for regulators in terms of ensuring periodic changes to the warnings, as well as studies that compare labelling policies across jurisdictions. For example, a recent study found that new text-based warnings, introduced in the United Kingdom in 2003, were considerably more likely to be noticed than Australian text-based warnings, which were only slightly smaller, but had been in place for more than eight years at the time of the survey (Bornstein, 1989). Ideally, labelling policies should be evaluated at similar post-implementation dates; at the least, differences in follow-up periods should be clearly noted and taken into account when interpreting findings.

There is preliminary evidence to suggest that not all measures of effectiveness decline at the same rate over time. "Proximal" measures of salience, such as noticing warnings, may erode more quickly than "distal" measures, such as reporting that health warnings motivate quitting and increase thoughts about the health risks of smoking (Hammond *et al.*, 2007a). It is even plausible that for some smokers the impact of health warnings could increase over time. For example, the cessation and telephone quitline information included in many health warnings may only become relevant to smokers as they contemplate quitting. In a population-based survey, however, the ebb and flow among individuals will balance out, and one would still anticipate decreases in measures of effectiveness over time.

Youth:

One policy-relevant question concerns the impact of warning labels in reducing youth uptake. Evaluating the impact of health warnings among youth during periods of smoking initiation requires a different conceptual approach. Given that the cigarette package serves as the medium for labelling policies, consumption levels may be positively associated with knowledge of the warning labels. In other words, individuals who smoke 20 cigarettes a day will be exposed to the warnings more frequently than individuals who smoke less than daily. Furthermore, "occasional" youth smokers are less likely to buy their own package, reducing the likelihood of exposure to warning labels, compared to more regular smokers who are more likely to buy their own package (Leatherdale, 2005). As a result, individuals who smoke more frequently are more likely to recall the content, location, and other aspects of labelling policies, a counter-intuitive association at first glance (Robinson & Killen, 1997).

A second issue concerns longitudinal studies that use measures of exposure or knowledge as predictors of future smoking behaviour among youth. During youth and young adulthood, the rate of smoking undergoes significant increases. As youth smoking behaviour increases, so too will their exposure to the package and their knowledge of the warnings. Thus, whereas a negative association

between exposure and future smoking behaviour may be expected for anti-smoking campaigns in other media, this is not the case for warning labels.

Failure to account for the somewhat counter-intuitive association between smoking and exposure to health warnings can result in misleading interpretations of data. For example, one study characterized an association between increased smoking and increased knowledge of health warnings as "paradoxical," and also found evidence that US health warnings were ineffective (Robinson & Killen, 1997). This may have been the case; however, without a comparison group, the authors had no way of knowing whether the increases in smoking behaviour were greater, less, or no different than they would have been if no warnings or more comprehensive warnings had been implemented. It may be, for example, that fewer youth initiated smoking than would have otherwise occurred without the health warnings. In fact, this was the pattern reported in a longitudinal study comparing changes in youth smoking in Canada and the USA following the introduction of graphic warning labels on Canadian packages. Smoking rates and knowledge of the warnings rose among Canadian youth as they aged; however, the increase in smoking was significantly less than among US adolescents and the increase in knowledge of the warnings considerably greater (Fong *et al.*, 2002). Overall, this study underscores the importance of suitable research designs and appropriate interpretations of the data when evaluating warning labels among youth.

Evaluation of individual messages & content:

Beyond the question of whether health warnings are generally effective, there is a growing body of research on the individual elements of a warning. These elements can broadly be categorized in terms of design and content components. To date, much of the research has focused on important design elements including the size, position, and use of pictures on the package (Strahan *et al.*, 2002). In contrast, relatively few studies have examined the content of individual messages.

Population-based surveys that compare labelling policies across time or jurisdictions are somewhat ill-suited to the task of evaluating individual warnings. Policies typically differ on more than one dimension, and policy changes typically involve increases in the size, number, position, and type of information presented in each warning. Evaluating individual components or messages becomes more complicated as the number of warnings and complexity of information increases; it is far easier to evaluate the effectiveness of a single warning through survey-based research than to evaluate the content of 16 individual warnings.

When assessing the impact of individual warnings, it is also important to consider that many health warnings are tailored to particular sub-groups of smokers. Warnings on the risks of smoking while pregnant, for example, have little relevance for older males. Thus, it is conceivable that some warnings may perform very well among sub-groups who comprise the target audience, but relatively poorly among the population as a whole. As a consequence, survey measures may need to be adapted and the findings may need to be stratified among relevant sub-groups. One might expect the tailoring of warnings to increase, as the use of picture-based warnings increase, along with the typical number of rotating warnings in a given jurisdiction.

In general, population-based surveys may be most appropriate for identifying the overall effectiveness of a set of health warnings. However, the task of evaluating the content of individual warnings is best suited to experimental or qualitative designs, in which the content and design can be systematically varied.

Geographic & cultural differences:

Very little research has examined potential geographic and cultural differences in the effectiveness of health warnings. Although the fundamental principles underlying the effectiveness of warnings are unlikely to vary across cultures and regions, the effectiveness of individual messages may indeed perform differently. First, smokers in different parts of the world have different levels of existing health knowledge. This has implications

for the type of messages to be included in warnings. For example, Australian smokers may have a relatively higher level of health literacy than smokers in other regions, which may account for the decision to include a warning for "peripheral vascular disease" on packages. Picture-based warnings may be particularly important in populations with lower literacy rates (CRÉATEC, 2003). In addition, the images used in one jurisdiction may not be equally effective in another. For example, several of the picture-based warnings that appear on Venezuelan and Uruguayan packages, and elsewhere, use symbols that may be culturally specific. Finally, similar sets of warnings may be more effective in areas where smokers have relatively little access to anti-smoking information from mass media or health professionals. Few of these issues have been addressed to date; however, they are likely to gain prominence as a growing number of jurisdictions in Asia, Africa, and the Middle East enhance their labelling policies to meet Article 11, and must rely on an evidence base that derives from relatively few Western and Latin American countries.

Evaluating the removal of information:

Unlike other labelling policies, restrictions on brand descriptors result in the removal, rather than the provision of, information. This presents a challenge to evaluation, particularly when the information being removed is used as a brand identifier. In the case of bans on the use of "light" and "mild," the terminology that was previously used to identify a class of products no longer exists. Smokers may retain the same misleading perceptions of these products after the terms have been prohibited, but survey measures can no longer refer to "light" or "mild" cigarettes in the same way as in the past. Therefore, survey measures must be designed so that the wording and meaning of questions remains constant before and after the removal of these terms. This creative challenge is only now being confronted by researchers with the recent advent of "light" and "mild" prohibitions. One approach, discussed later in this section, is to make the respondents' "own brand" the referent for questions.

Another implication of the "removal" of brand information is that the beliefs associated with "light" and "mild" cigarettes are likely to persist for some time after the descriptors disappear from packages. This situation is similar to advertising, promotion, and sponsorship bans; one should not expect beliefs to change immediately upon the implementation of the policy, but more gradually over time. Indeed, anecdotal evidence suggests that many retailers and consumers continue to use the terms "light" and "mild" well after their removal. Other packaging elements and aspect of cigarette design may also reinforce the same beliefs and perceptions as the "light" and "mild" descriptors. These considerations are important in terms of how the data are interpreted and how the "effectiveness" of light and mild policies is conceptualized.

Defining misleading descriptors:

There is widespread confusion among both consumers and many within the tobacco control community regarding several key terms relevant to labelling policy. Many fail to make the distinction between "light" and "mild" and "low tar." Whereas "light" and "mild" are terms used in the name of a brand, strictly speaking "low tar" refers to the emission levels under machine testing. Although there is a very strong correlation between the two (manufactures often attach "light" and "mild" descriptors to brands that generate lower tar levels under the ISO smoking machine), one can have a "light" cigarette that does not generate "low tar" levels and vice versa. Strictly speaking, in jurisdictions with bans, "light" and "mild" cigarettes do not exist, whereas "low tar" cigarettes do. To complicate matters further, the terms "light" and "mild" can also be used to refer to sensory properties of a cigarette. Thus, smokers may still retain the concept of a cigarette as "light" or "mild" even in the absence of a brand descriptor. Given the potential for confusion, survey measures should be explicit about the intended meaning of these terms and should avoid using them interchangeably. This becomes

apparent when measuring these concepts in jurisdictions where the "light" and "mild" brand descriptors have been removed.

Measures

This section provides an overview of the key constructs and individual measures that have been used to assess labelling policies. The constructs range from the extent to which labelling policies are noticed and processed, the extent to which they alter key beliefs (such as levels of health knowledge), to their impact upon downstream behavioural outcomes. These measures can be organised within a conceptual model, as illustrated in Figure 5.29. Other psychosocial variables, such as social norms and beliefs about the tobacco industry, could also be added to this model, but have been excluded in the interest of brevity. The following sub-section begins with a review of quantitative measures, followed by qualitative measures, and a brief discussion of the role of industry documents (Tables 5.27-5.39; see also Appendices 9 and 10).

Measures of labelling salience and processing:

Health warnings must be cognitively processed to be effective. The extent to which information is processed or elaborated upon has been shown to be the most important determinant of memory and attitude change in response to new information (Anderson, 1990). A number of measures have been developed to assess cognitive processing of health warnings as a general indicator of their salience. These measures range from more "shallow" measures of processing, such as a general awareness of warnings, to "deeper" measures of processing, including reading the warnings and thinking about them when they are not in sight (Borland & Hill, 1997a; Canadian Cancer Society, 2001; Hammond

Construct	Noticing Health Warnings
Measure	"In the last month, how often, if at all, have you noticed the warning labels on cigarette packs?" (Never, Rarely, Sometimes, Often, Very Often)
Sources	Hammond *et al*., 2006a; Hammond *et al*., 2007a
Validity	The time reference varies across different versions: some questions include no time reference ("How often do you notice…"), whereas others refer to the "last month" or "last three months." The response categories also vary and are often collapsed into a smaller number of categories in analysis. The basic question can also be asked within the context of noticing other forms of anti-tobacco media (e.g. "In the last 6 months, have you noticed advertising or information that talks about the dangers of smoking, or encourages quitting in any of the following places? (Yes, No to a list of 9 media channels, including on *cigarette packages*)).
Comments	Overall, a straightforward measure of the salience and processing of warnings that should be considered within the core set of variables to assess health warnings. As close to a "gold standard" in this domain as exists. Using the same wording to ask about salience of other media channels provides a useful comparative index for the salience of various health information channels. A recommended and essential measure for evaluating health warnings.

Table 5.27 Essential Measure of Labelling Salience and Processing

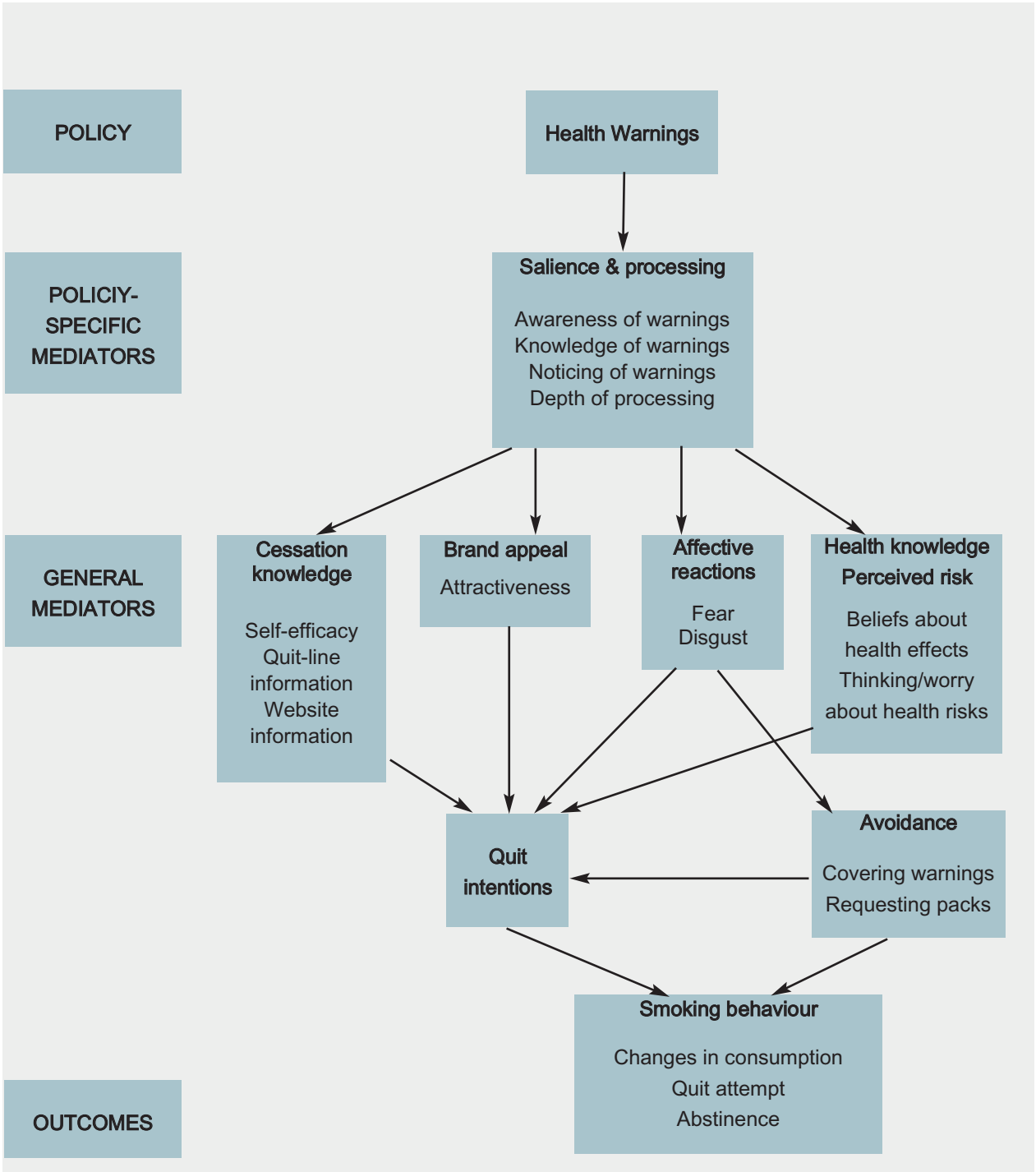


Figure 5.29 Conceptual framework for the evaluation of health warning policies

Construct	(a) General Awareness
Measure	"Have you seen health warnings on cigarette packages?" (Yes, No)
Sources	Borland & Hill, 1997a; Health Canada, 2005
Validity	Good face validity; associated with policy strength.
Variations	Response categories are consistent across measures. Alternative wordings include: "Are you aware of health warnings on cigarette packages?" and "Are there health warnings on packages?" Some questions refer specifically to the release of new warnings. For instance, "Are you aware of any recent changes to health warnings on cigarette packs?" and "Have you noticed any changes to the health warnings on cigarette packages since [date]?"
Comments	Provides an overall measure of general awareness. Limited value in examining changes and comparing across jurisdictions, given almost universal awareness among smokers. Most useful for examining policy implementation and rollout when the question makes reference to "new" warnings, or in jurisdictions with very weak health warnings and no previous research. Overall, an informative measure, but only recommended under these circumstances.
Construct	**(b) Reading/Looking Closely at Health Warnings**
Measure	"In the last month, how often, if at all, have you read or looked closely at the warning labels on cigarette packs?" (Never, Rarely, Sometimes, Often, Very Often)
Sources	Hammond *et al*., 2006a; Hammond *et al*., 2007a
Validity	Face validity; good convergent validity with other measures; good predictive validity for strength of policy and motivation to quit smoking.
Variations	The time reference varies across different versions. Also, some versions refer to *reading*, other versions use broader language, such as *looking closely,* and some versions include both terms. *Looking closely* may be more appropriate for pictorial warnings.
Comments	Strong correlation with noticing, but conceptualized as a "deeper" measure of processing. Overall, a recommended and important, but not essential, measure of salience and processing that may be particularly relevant for textual aspects of warnings.
Construct	**(c) Discussing the Health Warnings With Others**
Measure	"In the last month, how often, if at all, have you talked about the health warning with others?" (Never, Rarely, Sometimes, Often, Very Often)
Source	Hammond *et al*., 2003

Table 5.28 Additional Measures of Labelling Salience and Processing

Validity	Good face validity, convergent validity, and predictive validity for motivation to quit and future smoking behaviour when included as part of a composite measure.
Variations	Variations of this measures use slightly different terms, including *discussed* and *mentioned* rather than *talked about*, as well as different response options, such as Never, Rarely, Sometimes, Frequently.
Comments	These measures provide a "deeper" measure of processing for labels and may be useful for comprehensive evaluations of labelling policies. Recommended, but not essential.
Construct	**(d) Thinking About Health Warnings**
Measure	"In the last month, how often have you thought about what the health warnings have to say?" (Never, Rarely, Sometimes, Often, All the time)
Sources	Canadian Cancer Society, 2001; Hammond *et al.,* 2003; Christie & Etter, 2004
Validity	Good face validity, convergent validity, and predictive validity for motivation to quit and future smoking behaviour when included as part of a composite measure.
Variations	"In the last month, have you ever thought about the warning labels or what they had to say when a cigarette pack wasn't in sight?" This variation of the measure requires a higher threshold of processing than the items above.
Comments	These measures provide a "deeper" measure of processing for labels and may be useful for comprehensive evaluations of labelling policies. Recommended, but not essential.

Table 5.28 Additional measures of labelling salience and processing

et al., 2003; Christie & Etter, 2004; Hammond *et al.*, 2004a; Health Canada, 2005; Koval *et al.*, 2005; Hammond *et al.,* 2006a; Hammond *et al.,* 2007a).

Measures of general "awareness" are typically endorsed by a vast majority of respondents, including non-smokers, regardless of the type of warning level. These questions are often used to examine the implementation, or "roll-out," of new package warnings following a change in policy. This information is critical for any population-based survey conducted shortly after the implementation of a new policy, given the uncertainty regarding when health warnings begin appearing on packages.

In contrast to general measures of awareness, the extent to which smokers notice, read, and think about the warnings appears to be highly dependent on the size, type, and location of the warning (Borland & Hill, 1997a; Health Canada, 2005; Hammond *et al.*, 2007a). These measures of processing are also subject to the implementation date. Several studies have used measures of processing collected from the same population over time and can be used to measure the "wear-out" (i.e. decrease in the salience of the warning labels) of health warnings (Health Canada, 2005; Hammond *et al.*, 2007a). Additional data of this type may help to answer the question as to whether the rate of decline among measures of salience is associated with design features, such as the size of warnings and the use of pictures. In most cases, these measures have been analyzed as individual items, although in one case a *depth of processing* scale was developed and tested (Hammond *et al.,* 2003). In that instance, nine items were used to create a scale to

Construct	Health Warnings - Eye Tracking
Measure	Participants wore eye-tracking equipment and viewed US cigarette advertisements with health warnings.
Sources	Fischer *et al.*,1989b; Krugman *et al.*, 1994
Validity	Good predictive validity for recall and recognition of health warnings
Variations	Viewing time serves as another measure of attention, where warnings are flashed on a screen and the amount of time is recorded (Peters *et al.*, 2007).
Comments	Eye tracking measures can help to identify the most salient design aspects of warning labels and serve as an objective measure of attention; however, these measures are limited to "laboratory" based research designs.

Table 5.29 Physiological Measures of Salience and Processing

measure cognitive processing labelled as "depth of processing." Responses to each of the nine items were rated using a 5-point Likert-type format going from "not at all/never" to "all the time/a lot" and values added to create an index. Examples of items included were "How carefully have you ever read the messages on the outside of a cigarette package?" and "How often have you thought about what messages on the inside of packages have to say?"

Although the wording of items is relatively similar across surveys, different time periods are used in both the question and the response option in many cases. For example, whereas some "noticing" questions refer to the past month, others refer to the past three months, or use no time reference at all (Tables 5.27 and 5.28). Nevertheless, findings from the same population are relatively similar across different question wordings (Canadian Cancer Society, 2001; Hammond *et al.*, 2004; Health Canada, 2005; Hammond *et al.*, 2007a).

Contents & emissions:

Several studies have assessed the extent to which smokers process emission information printed on the side of packages. These measures mirror the processing items used to gauge health warnings, although a more limited set has been used. Both studies of which we are aware, indicate that smokers are less likely to read or look at emission information than health warnings on the face of packages (Thompson *et al.*, 2006). More generally, it is unclear whether salience and processing type measures are as informative for emission labelling policies as for health warning policies. Unlike health warnings, which typically include a number of rotating health warnings, emission labelling is consistent across packages for a given brand. As a result, there may be little reason for smokers to read or attend to this information on a regular basis. As a consequence, we have not recommended a specific measure in this section.

Physiological measures of salience and processing:

Physiological measures have been used in conjunction with survey measures to quantify attention to and processing of health warnings. These measures have an advantage in that they are more "objective" given that they do not rely on self-reporting. In several cases, they have been used to compare the salience of warnings with package design or within the context of a tobacco advertisement. For

Construct	(a) Health Warnings - Location
Measure	Without looking at a cigarette package, where on the pack are the warnings or messages located?" (Open ended)
Source	Hammond *et al.*, 2003
Validity	Good face validity.
Variations	The same question has been asked without the prefix ("Without looking at a cigarette package…"), as well as with a diagram in self-completed surveys.
Comments	Useful measures for identifying basic knowledge about health warnings; however, it becomes complicated in jurisdictions with warnings on the inside and outside of packages. Unclear how emission and contents information should be treated, especially when provided by industry.
Construct	**(b) Health Warnings – Content**
Measure	"Without looking at a cigarette package, what specific health warning messages can you remember seeing on cigarette packages in Canada?" (Open ended)
Source	Health Canada, 2005
Validity	Good face validity.
Variations	The same question has been asked without the prefix, which is typically included in telephone surveys to ensure the participant is not looking at the package during the call.
Comments	Useful measures for identifying basic knowledge about health warnings and, potentially, for identifying individual messages that are particularly salient. However, this measure will be difficult to answer in jurisdictions with comprehensive health warnings, including multiple warnings on different areas of the package.

Table 5.30 Measures of Knowledge of Health Warnings

example, eye movements during exposure to an ad have been used as physiological indicators of attention to tobacco warnings. These measures that are directly linked to cognitive processing have been useful to investigate the relationship between visual attention and a more traditional communication measure (Krugman *et al.*, 1994) (Table 5.29).

Knowledge of health warnings

Items assessing smokers' knowledge of health warnings are among commonly used survey measures. Knowledge questions have been asked using unprompted recall (e.g. "Where are the warnings located?"), as well as using recognition tasks (e.g. "Please tell me which of the following warnings appear on cigarette packages…") (Table 5.30) (Hill, 1988; Richards *et al.*, 1989; Rootman *et al.*, 1995;

Construct	Emission Side Panel – Content
Measure	"Without looking at a cigarette package, can you name any chemicals or substances that are currently listed on cigarette packages in [country]?" (Open ended)
Source	Health Canada, 2003
Validity	Face validity.
Variations	A common alternative is to ask about the quantitative level of specific emissions, such as tar "Without looking at a pack, can you tell me the tar level of your cigarettes?"
Comments	This measure examines basic recall of emission information printed on packages, and is often compared against "objective" data collected from other sources in order to evaluate accuracy of self-report recall. This measure should be interpreted alongside measures on the comprehension and use of this information (described later).

Table 5.31 Measures of Knowledge of Constituents and Emissions

Construct	Health Warnings – Affective Reactions
Measure	"Have you experienced any fear as a result of the health warnings?" (Not at all / A little / A lot)
Source	Hammond *et al.*, 2004a
Validity	Good face validity; good predictive validity for future smoking behaviour.
Variations	Alternatives have used more comprehensive scales and asked about different affective reactions, including disgust and anger (Peters *et al.*, 2007).
Comments	Affective reactions have been evaluated to a greater extent in qualitative evaluations of warning labels; however, survey-based measures may be a key mediator of downstream measures of impact.

Table 5.32 Measures of Affective Reactions to Health Warnings

Borland & Hill, 1997a; Borland & Hill, 1997b; Robinson & Killen, 1997; Hammond *et al.*, 2003; Brown *et al.*, 2005; Health Canada, 2005; O'Hegarty *et al.,* 2006; Thompson *et al.*, 2006). Measures of unprompted recall for warning label content can be used to identify which individual warnings may be most effective. In jurisdictions with a large number of warnings, this task can be particularly helpful.

Many of these measures have been assessed among the general population, including among nonsmokers. Except for the few questions that refer to a respondent's "own" cigarette package, most measures of awareness and knowledge appear to work equally well among nonsmokers. Indeed, nonsmokers have been found to have

Construct	Health Warnings – Avoidance
Measure	"In the last month, have you made any effort to avoid looking at or thinking about the warning labels?" (Yes, No)
Sources	Hammond *et al*., 2004a; International Tobacco Control Policy Evaluation Survey (The ITC Project)
Validity	Good face validity; good predictive validity for future smoking behaviour.
Variations	Several follow-up questions may be asked of those who respond "yes" to the initial question, above. For example, "Have you made any effort to avoid the warnings by: (1) Covering the warnings up? (2) Keeping the pack out of sight? (3) Using a cigarette case or some other pack? (4) By not buying packs with particular labels?" (Yes, No to each question)
Comments	These measures can indicate the prevalence of avoidance behaviours and whether they reduce the effectiveness of warnings. The follow-up questions are only necessary for in-depth exploration of avoidance.

Table 5.33 Measures of Avoidance

surprisingly high levels of awareness and recall for prominent health warnings and picture-based warnings in particular (Health Canada, 2005). However, both recall and recognition of particular messages has been shown to be highly dependent on the complexity of the health warning and its implementation date. For example, virtually all Canadian smokers are aware of the health warnings on packages, although we are unaware of any research indicating that smokers have correctly been able to identify all 16 health warnings that appear on packages.

Analyses must take into account the consumption level when assessing knowledge of health warnings. Given the inevitable link between heaviness of smoking and viewing the warning labels, knowledge is likely to be greater among heavier smokers. This association is likely to be more pronounced within samples that include a broad range of smokers, and are likely to be greatest in studies that compare regular smokers with occasional or nonsmokers. The association between consumption and knowledge is also likely to be stronger in jurisdictions with a greater number and complexity of warnings. For example, packages in Canada carry information on the side panel, one of 16 health warnings on the outside of packages, and one of 16 additional warnings on the inside of packages. In such cases, a greater number of exposures will be required to recall various aspects of the warnings.

There are several limitations with measures of knowledge. First, when asking about the location of health warnings, one issue is whether respondents consider emission information, which may be printed on the sides of the package, as a health warning. Canadian data suggests that some smokers are aware of this information, but fail to cite it as a location. Second, in telephone or web-based surveys, some participants may have a pack visible as they respond to the survey. As a result, some measures explicitly ask smokers not to look at the package to avoid this situation to the extent possible. Third, measures of knowledge can often be difficult to compare across labelling policies. For example, smokers from the USA, where a total of four different text warnings appear on packages, have a much greater

Construct	(a) Health Warnings – Believability/Credibility
Measure	"Overall, do you believe the health warning message(s)?" (Not at all, A little, A lot)
Source	Health Canada Youth Smoking Survey (http://www.hc-sc.gc.ca/hl-vs/pubs/tobac-tabac/yss-etj-2002/index-eng.php)
Validity	Good face validity.
Variations	Other alternatives refer to the *accuracy, trustworthiness, credibility, believability* and *true/false* nature of the warnings or the *importance* of information (Cecil *et al.*, 1996; Borland & Hill, 1997a; Canadian Cancer Society, 2001; Hammond *et al,* 2004a; Brown *et al.*, 2005; Health Canada, 2005; O'Hegarty *et al.*, 2006). Some surveys have also included more comprehensive, but also more time consuming, scales involving numerous items.
Comments	A useful, brief measure to examine credibility of message content. The measure can be used to examine whether different design and content features change the believability of information among smokers. This question can be asked of individual health messages, such as in qualitative or experimental research, or to refer to a set of warnings, as is common in population-based surveys. Note that responses to this item will also reflect denial, self-exempting beliefs, etc.
Construct	**(b) Health Warnings – Public Opinion/Support**
Measure	"Do you approve of the health warnings on cigarette packages?" (Yes, No)
Source	Borland & Hill, 1997a
Validity	Good face validity.
Variations	Other alternatives include measures of agreement with the warnings and references to *appropriateness or desire for more information* (Canadian Cancer Society, 2001; Hammond *et al.*, 2004a; Brown *et al.*, 2005., Health Canada, 2005)
Comments	This measure is a combination of previously administered questions and has yet to be administered exactly as worded. Though measures of public support or approval may be less important as a measure of effectiveness, they are a critical measure for regulators and policy makers, and for demonstrating support for more comprehensive policies.

Table 5.34 Measures of Credibility and Public Support

likelihood of correctly identifying all the messages than smokers in the United Kingdom where 16 different text messages appear on packages. The same issue arises in pre-post studies of a new labelling policy. For example, when Canada revised its labelling policy in 2000 to include pictures, the number of individual messages doubled from eight to 16 (not counting 16 additional messages that appeared on the inside of packages). In such cases, neither the total number nor the proportion of messages correctly identified, provide a suitable basis for comparing policies given that the denominator is different. Moreover, it is both time consuming

Construct	Health Warnings – Thinking About Health Risks
Measure	"To what extent, if at all, do the warning labels make you think about the health risks of smoking?" (Not at all, A little, A lot)
Source	Hammond *et al.*, 2007a
Validity	Good face validity; good convergent validity; associated with strength of policy.
Variations	Similar questions ask about the extent to which warnings affect the level of concern or *worry* about health risks.
Comments	A key mediator of the effectiveness of health warnings. This should be considered among the essential measures.

Table 5.35 Measures of Health Knowledge and Perceived Risk

Construct	Emissions Information – Comprehension
Measure	"If you were to look for a safer or less harmful cigarette, would you use information about the amounts of chemicals listed on the cigarette packs to help you find a less harmful brand?" (Yes, Maybe, No)
Sources	Gori, 1990; Health Canada, 2003
Validity	Good face validity.
Variations	Similar questions ask smokers to compare different tar levels of cigarettes in terms of delivery and health risks.
Comments	A critical measure to evaluate emission policies that include quantitative emission levels. The question can also be used to refer to specific emissions, such as tar or nicotine. This measure is essential in any survey that also asks about recall or awareness of emission numbers on packages. The current wording can be used to refer both to descriptive (i.e. text-based) and quantitative emission information.

Table 5.36 Measurement of Comprehension of Emissions Information

and awkward to prompt survey respondents for 16 different warnings.

Finally, some knowledge measures may not work across all survey modalities. For example, Krugman and Robinson presented participants with diagrams of various warnings in a recognition task (Krugman *et al.*, 1994; Robinson & Killen, 1997). Any such measures, which require visual information to be presented to participants, must be administered either face-to-face or using web-based modalities.

Construct	(a) Light / Mild Descriptors – Comparative Risk
Measure	"Light cigarettes are less harmful than regular cigarettes." (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly Disagree)
Source	The ITC Project
Validity	Good face validity; good convergent validity (Borland *et al.*, 2004).
Variations	This question can be adapted to refer to other descriptors, such as *mild* or *smooth*. In some cases, the terms *light* and *mild* are used in the same question. Alternatives ask smokers about differences in the "tar" or "nicotine" of *light* versus *regular* cigarettes (Smokers of light cigarettes take in less tar than smokers of regular cigarettes). These measures have been widely used, but require a basic familiarity with tar and nicotine, which may not exist in all smokers in some jurisdictions (Kozlowski *et al.*, 1998b; Shiffman *et al.*, 2001; Hamilton *et al.*, 2004). Other alternatives have asked smokers to report the number of *light* cigarettes that would need to be smoked to equal the harm from 10 regular cigarettes; however, this approach requires a level of numerical literacy beyond the capacity of smokers in many jurisdictions (Kozlowski *et al.*, 2000; Shiffman *et al.,* 2001).
Comments:	This is an essential construct, although there is no single "gold standard" question for its measurement. The recommended measure has been selected because is it the most direct and may be most appropriate for smokers in low- and middle-income countries. Nevertheless, the question may need to be preceded by a general awareness questions (e.g. "Have you ever heard of *light* cigarettes?") in some markets or rural areas. There are also issues with the interpretation of this measure in jurisdictions where *light* and *mild* descriptors have been prohibited.
Construct	**(b) Light/Mild Descriptors – Addiction**
Measure	"Light cigarettes are less addictive than regular cigarettes." (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly Disagree)
Source	The ITC Project
Validity	Good face validity; good convergent validity (Borland *et al.*, 2004).
Variations	This question can be adapted to refer to other descriptors, such as mild or smooth. In some cases, the terms *light* and *mild* are used in the same question.
Comments	A straightforward question with the same format and response options as above. A recommended question to address perceptions of *light* and *mild* cigarettes, but not as essential as the comparative risk question, above.

Table 5.37 Measures of Light, Mild, and Brand Descriptors

Construct	Brand Appeal – Health Warnings
Measure	"Do you think the new warnings make cigarettes packages look less attractive, more attractive, or has it made no difference to their attractiveness?" (Not at all, A little, A lot) "How often have you put your cigarette package away because you didn't want others to see the warning on the package? Have you done this?" (Never, Sometimes, Often)
Source	Canadian Cancer Society, 2001
Validity	Face validity.
Variations	Alternatives refer to quality of advertisements with and without warnings, whether youth would want to "use" the product, intentions to purchase the product in the future, and a measure of perceived economic values of brands (Hyland & Birrell, 1979; Brubaker & Mitby, 1990; Canadian Cancer Society, 2001; Willemsen *et al.*, 2002; Thrasher *et al.*, 2007). "Attractiveness" scales have also been used (Loken & Howard-Pitney, 1988).
Comments	These measures provide a straightforward evaluation of whether health warnings have altered the general appeal of packaging. The second of the two measures has a higher threshold and represents a more distal measure of appeal, which may also tap into social norms. Both of the measures are recommended for surveys that wish to provide a comprehensive evaluation of warnings, but are not essential.

Table 5.38 Measures of Brand Appeal

Constituents & emissions:

A number of studies have examined whether smokers can recall the emission information commonly printed on the side panel of cigarette packages (Table 5.31) (Chapman, 1986; Cohen, 1996b; Health Canada, 2003; O'Connor *et al.*, 2006c). These items typically ask participants to name the emissions printed on packages using unprompted recall tasks, or ask them to report the number associated with a particular emission (usually "tar"). The data indicates that many smokers have a general awareness that tar and nicotine numbers may be printed on the package, but few are able to recall the tar or nicotine levels printed on their usual brand of cigarettes. To our knowledge, no measures have been developed to measure smokers' knowledge of tobacco contents.

Affective reactions to health warnings:

Research in the field of health communication indicates that messages with emotionally arousing content are more likely to be noticed and processed by smokers (Witte & Allen, 2000). Strong emotional responses to messages are also associated with greater behaviour change when supportive or "efficacy" related information is also presented. To date, several studies have used measures of affective reactions to assess the impact of warnings labels (Environics Research Group, 2000; Elliot & Shanahan Research, 2002; Environics Research Group, 2003; Hammond *et al.,* 2004a; Health Canada, 2006; Peters *et al.*, 2007). These measures are common in qualitative evaluations of individual warning labels and have been particularly influential in development of picture-based warnings in

Construct	(a) Changes in Foregoing – Health Warnings
Measure	"In the last month, have the warning labels stopped you from having a cigarette when you were about to smoke one?" (Never, Once, A few times, Many times)
Sources	Borland & Hill, 1997a; Hammond *et al.*, 2007a
Validity	Good face validity; convergent validity; associated with strength of policy.
Variations	Similar measures have referred to not smoking when *tempted.*
Comments	This question has a lower "threshold" than other measures that assess the behavioural effects of health warnings.
Construct	**(b) Reductions in Smoking – Health Warnings**
Measure	"Are you smoking any less or more as a result of the new warnings, or are you still smoking the same amount?" (Less, Same amount, No difference)
Source	Hammond *et al.*, 2007a
Validity	Good face validity; convergent validity; associated with strength of policy.
Variations	Similar measures have referred to not smoking when *tempted.*
Comments	The wording "as a result of the warnings" needs to be emphasized when asking this question. This item is not intended to provide a precise measure of changes in consumption as a result of the warnings; changes in consumption happen in response to a wide range of related factors. However, this question does provide a good general measure of the extent to which smokers have been affected by the warnings.
Construct	**(c) Likelihood/Motivations to Quit**
Measure	"To what extent, if at all, do the warning labels on cigarette packs make you more likely to quit smoking?" (Not at all, A little, A lot)
Source	Hammond *et al.*, 2007a
Validity	Good face validity; convergent validity.
Variations	Alternatives refer to *motivations to quit* and *thinking about quitting*, with some differences between response categories.
Comments	The recommended wording refers directly to the likelihood of quitting smoking, which is somewhat broader than motivation alone. In practice, however, there appears to be few differences with regards to how these measures perform in practice given the consistency of findings from similar samples. The question has the potential to provide a very good summary measure of the self-reported impact of health warnings and should be considered within the core set of items to evaluate labelling policy.

Table 5.39 Measures of Behavioural Outcomes

Construct	(d) Quit Attempts & Abstinence
Measure	"To what extent, if at all, were the following reasons for your current quit attempt…warning labels on cigarette packages?" (Not at all, A little, A lot)
Source	The ITC Project
Validity	Good face and convergent validity.
Variations	Alternatives have also asked about the effect of the warnings on staying quit in the future. This question can be asked as part of a list of reasons for quitting, which provides some useful context on the relative influence of other potential influences on quitting.
Comments	Retrospective measures, such as this, should be interpreted with caution given that they are subject to recall biases, particularly as the time since the quit date increases. In addition, smokers often cite a number of complementary reasons for quitting and endorsement of this item does not mean that the quit attempt is solely attributable to health warnings.

Table 5.39 Measures of Behavioural Outcomes

several jurisdictions. Measures of negative emotions, including fear and disgust, have also been used in population-based surveys and shown to predict future cessation-related behaviour (Table 5.32). Overall, measures of emotion have considerable promise as a proximal measure of effectiveness which can be used in both qualitative and quantitative research.

Avoidance:

Warnings that result in unpleasant emotions may lead some smokers to avoid the warnings. Indeed, several studies indicate that a considerable portion of smokers make some attempt to avoid the warnings, including covering or hiding the warnings, using another case, or requesting different packs to avoid particular warnings. In some jurisdictions, tobacco manufacturers have been accused of marketing covers specifically intended to cover picture-based warnings, prompting calls for regulatory bans on the sale of such covers (Table 5.33) (Wilson *et al.,* 2006).

Although avoidance behaviours may be undesirable to some extent, these examples of fear control behaviour do not necessarily reflect an adverse outcome or inherent weakness of package warnings. Research has demonstrated that avoidant behaviours and attempts at thought suppression often have the opposite effect of increasing the presence of the unwanted thoughts (Wegner, 1994). In the context of the warning labels, avoidant behaviour might be more reasonably interpreted as a measure of effectiveness. Indeed, if the warnings were ineffective in communicating the threatening consequences of smoking there would be no reason to avoid them. Furthermore, one study found that smokers who attempted to avoid the warnings were no less likely to see the warnings, think about them, or engage in cessation behaviour at a 3-month follow-up (Hammond *et al.*, 2004a).

Credibility & public support:

In order to be effective, the health information presented in warnings must be credible. The credibility of warnings relates not only to the health information contained in a warning, but also to its design and source or attribution. Some have even speculated that there may be a trade-off between the vividness of the information in health warnings and its credibility among smokers. In others words, if

pictures and text become too striking or graphic, smokers may begin to question the accuracy of the information and become more resistant to the messages.

Although some validated scales have been used to evaluate the believability of health warnings (e.g. Beltramini, 1988; Loken & Howard-Pitney, 1988; Cecil *et al.*, 1996), many studies have used single questions with face validity (Borland & Hill, 1997a; Canadian Cancer Society, 2001; Hammond *et al.*, 2004a; Brown *et al.,* 2005; Health Canada, 2005; O'Hegarty et al., 2006; Peters *et al.*, 2007). Together, the findings suggest that health warnings represent a credible source of information, particularly when attributed to a well-respected department of health, or a well-respected non-governmental authority, such as a cancer society (Guttman & Peleg, 2003; Health Canada, 2003; BRC Marketing & Social Research, 2004). The levels of credibility do not appear to be associated with the type or design of warning labels; just like for text-based warnings, smokers report high levels of believability for graphic picture-based warnings as well.

Several studies have also sought to assess general measures of public support for health warnings (Borland & Hill, 1997b; Brown *et al.*, 2005; Hammond *et al.,* 2004a; O'Hegarty *et al.*, 2006). To our knowledge, two items have been developed to examine support among smokers for emission labelling ("Overall, do you believe the health warning message(s)?" and "Do you approve of the health warnings on cigarette packages?") (Health Canada, 2001; Health Canada, 2003). Public opinion data may be particularly effective for policy makers in gauging political support for new or existing labelling polices (Table 5.34).

Health knowledge & perceived risk

The primary objective of cigarette warning labels is to communicate the health effects from smoking. Thus, measures of health knowledge and perceived risk represent critical components in any evaluation of health warnings (Table 5.35). To date, studies have taken two main approaches to measuring the impact of warnings on health knowledge. One approach is to ask participants to self-report whether health warnings have changed the extent or frequency with which they think or worry about the health effects of smoking. Alternatively, some studies have assessed health knowledge directly and examined changes over time or across jurisdictions in levels of knowledge. Given the number of health effects caused by smoking, we are unaware of any study that has attempted to measure a complete list. However, studies typically measure beliefs about a range of specific health effects to determine knowledge levels. Some studies have included "bogus" health effects in the list in order to identify response bias. Most lists include "major" health effects, such as lung cancer and heart disease, as well as health effects on nonsmokers, and lesser-known health effects. Including lesser-known health effects can be particularly effective in attributing changes in knowledge to specific labelling policies. Ideally, longitudinal studies, assessing changes in health knowledge, would also select the health effects based upon the effects that are targeted in the warnings. In other words, studies should include health effects that: a) are already included on packages at baseline (before policy change) and will remain on packages at follow-up; b) health effects that are not on packages at baseline, but will appear at follow-up; and c) health effects that are not on packages at either baseline or follow-up. This type of design provides a measure of specificity with respect to changes in labelling policies.

A similar approach has been taken with respect to emission information. At least one study has examined whether knowledge of the emissions in tobacco smoke is higher in jurisdictions where they are printed on the package (Hammond *et al.*, 2006a). As with health effects, lists should include emissions that are, and are not, printed on packages, in order to examine the specificity of the effect.

Overall, research conducted to date suggests that increases in the size, number, and content of warnings are associated with greater thoughts about the health risks of smoking (Health Canada, 2005; Hammond *et al.*, 2007a).

More prominent warnings have also been associated with increased knowledge for specific health effects (Borland & Hill,1997a; Hammond, 2006a). Most of these findings derive from population-based surveys, although one study reported significantly higher beliefs about health effects following presentation of graphic versus text warnings within an experimental setting (O'Hegarty *et al.*, 2006).

Constituents & emissions:

A number of studies have sought to examine the extent to which smokers understand and interpret quantitative cigarette emission information (Table 5.36). These studies ask smokers to report either the "meaning" of the numbers, or the extent to which the numbers translate into differences in exposure from different brands (Gori, 1990; Cohen, 1996a; Health Canada, 2003; Thompson *et al.*, 2006). Other questions ask smokers to predict the health consequences of different tar levels, without explicit reference to labelling policies (Gori, 1990; Cohen, 1996a). Indeed, a number of studies on this topic were conducted in the USA, where there are no mandatory requirements to print emission levels on packages, they appear on packages less than 15% of the time, and are at the discretion of the manufacturer (Davis *et al.,* 1990).

Regardless of the jurisdiction or the labelling policy, the findings indicate that smokers have very little or no understanding of the meaning of the emission levels, although a substantial proportion associate health benefits with lower numbers. This type of data is critical to place measures of knowledge into context; prominent labelling that succeeds in increasing knowledge of emission levels is of little value if smokers do not understand the meaning of these numbers. Indeed, the data appear to indicate that communicating quantitative emission levels promotes erroneous perceptions about exposure levels and health risks that can be expected from different products. In general, this set of findings underscores the importance of assessing more than basic recall of information (Figure 5.30).

Light & mild descriptors:

A variety of surveys have examined perceptions of "light" and "mild" brand descriptors (Kozlowski *et al.*, 1998b; Kozlowski *et al.*, 2000; Ashley *et al.*, 2001; Shiffman *et al.*, 2001; Etter *et al.*, 2003c; Borland *et al.*, 2004; Hamilton *et al.*, 2004). Both quantitative and descriptive measures have been used to assess the health consequences of smoking "light/mild" cigarettes. Several studies have asked smokers how many light or ultra-light cigarettes would need to be smoked to inhale the equivalent level of tar as regular cigarettes. Some of these measures used "10 cigarettes" as a reference point, whereas others were open-ended. Smokers have also been asked to make comparisons between "light/-ultra-light" and "regular" brands using qualitative or descriptive categories to describe exposure levels and health risks. These qualitative response categories have also been used to compare perceived sensory properties and addiction levels of "light/mild" cigarettes compared to "regular" brands. At least one study combined items to create a "sensory" index and a "health effects" index (Shiffman *et al.*, 2001). Overall, both qualitative and quantitative measures appear to yield similar findings, and indicate that a substantial proportion of smokers perceive health benefits from cigarettes with "light" and "mild" descriptors (Table 5.37).

At least one study, the International Tobacco Control Policy Evaluation Survey (the ITC Project) (Borland *et al.*, 2004), has adopted an alternative approach to comparative estimates. Rather than asking smokers to compare "regular" and "light" cigarettes, participants were asked to compare their "usual" brand with regular cigarettes (e.g. "Do you think that the brand you usually smoke, [current brand], might be a little less harmful, no different, or a little more harmful, compared to other cigarette brands?"). Separate items were used to collect the name, descriptors, and relevant attributes of participants' "usual" brand. This approach has the benefit of personalizing the question, and is particularly useful to implement following the removal of "light" and "mild" terms, at which point questions with direct reference to "light" and

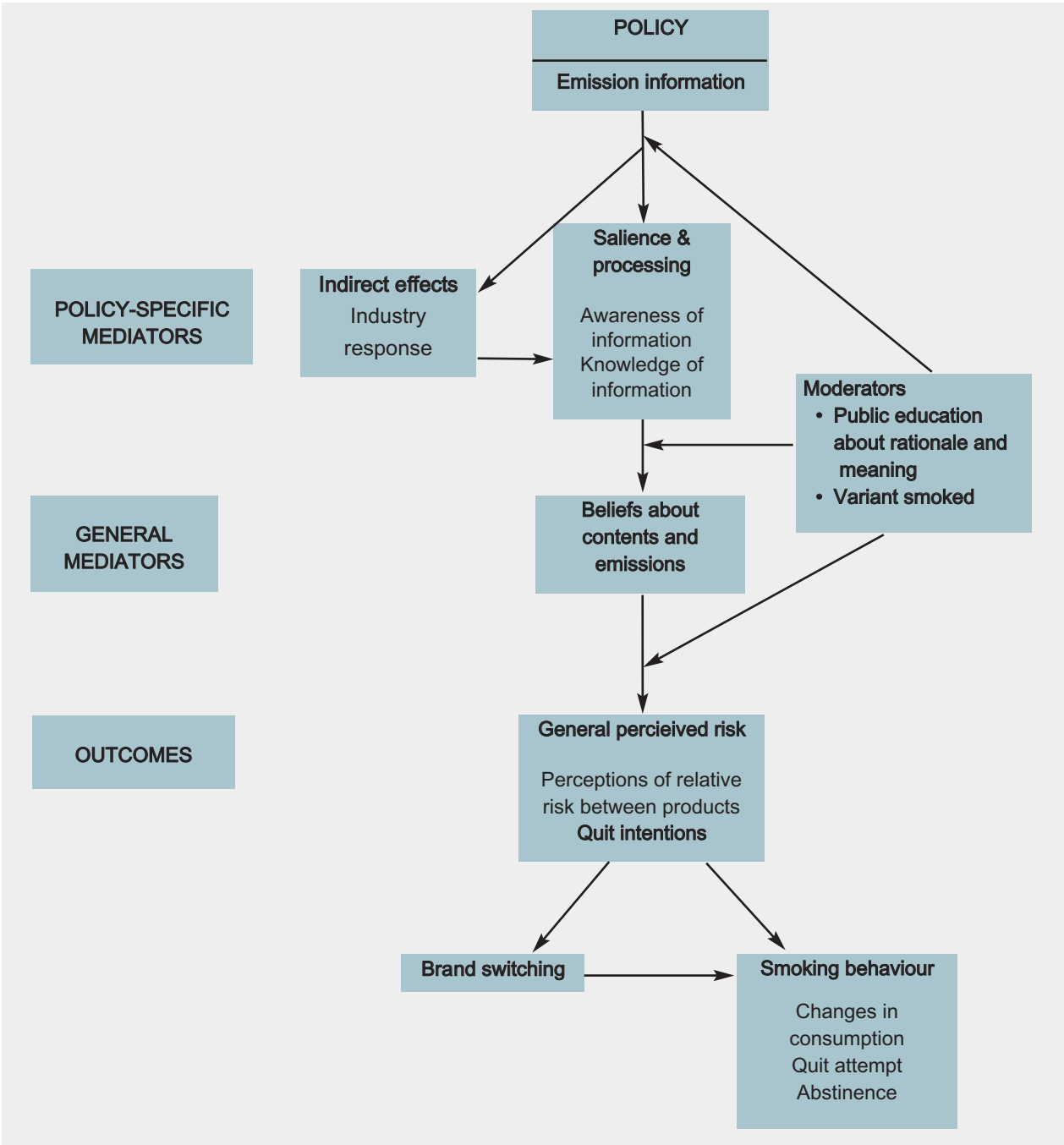


Figure 5.30 Conceptual Framework for the Evaluation of Emissions and Contents Labelling Policies

"mild" cigarettes become awkward and confusing. The question also has a broader frame of reference. This is an advantage in the sense that it captures the effect of other potential misleading descriptors or product elements. The disadvantage is that information on the respondents' own brand must also be available (see Section 5.4), and there is less specificity with respect to the brand elements that underlie differences in perceptions of risk. A similar conceptual approach has recently been taken with respect to evaluating print advertisements. Rather than asking smokers to compare the risks implied by the expressions "light" versus "regular" cigarettes, respondents were asked to rate the perceived risk to their health derived from advertisements for different products, and the ratings for advertisements of "light" versus "regular" cigarettes were compared (Hamilton *et al*., 2004). In most cases, follow-up questions may be necessary to identify which specific elements underlie perceptions of reduced harm.

Descriptors other than "light/-mild" are likely to receive increased attention in the coming years, particularly within jurisdictions where "light/mild" terms have already been prohibited. To our knowledge, only one study has developed measures to evaluate health perceptions based on other brand descriptors, including the words "smooth" and "ultra" (Thompson *et al*., 2006). Furthermore, studies with a focus upon brand descriptors in jurisdictions that have banned "light" and "mild," may wish to consider additional measures that examine the substitution of terms in their place. Market-based research, such as cataloguing the information printed on packages, can provide "objective" data on the substitution of terms which may be helpful in interpreting self-reported brand data (see Section 5.4).

Largely, the selection of measures in this area may depend upon the current state of policy more so than other areas (Figure 5.31).

Brand appeal

Health warnings target psychosocial variables other than perceived risk and health knowledge. More recent labelling policies include themes of addiction, industry manipulation, aesthetic costs, financial costs, and cessation beliefs, among others. A range of psychosocial measures have been developed to assess each of these constructs, although these measures have rarely been used to evaluate warning labels.

One area that has been explored is the impact of health warnings on measures of brand appeal (Table 5.38). In theory, replacing brand imagery with health warnings has the potential to change perceptions of the cigarettes and packaging. To date, the limited findings in this area appear to support this hypothesis, although it has yet to be explored in much depth with respect to warnings on packages (Hyland & Birrell, 1979; Loken & Howard-Pitney, 1988; Brubaker & Mitby, 1990; Hammond *et al*., 2004b; Thrasher *et al*., 2007). Future research might also explore whether larger graphic health warnings undermine the visual appeal of cigarette displays at retail outlets.

Behavioural outcomes

There are several approaches to predicting "downstream" cessation-related outcomes from health models. As with health effects, some studies have used measures of processing and knowledge of the warnings, and modelled their effects on motivation to quit and patterns of smoking behaviour (see Section 3.1 for measures of tobacco use and Section 3.2 for psychosocial outcomes). This has produced significant findings in longitudinal studies to date (Hammond *et al.,* 2003). However, this approach is somewhat limited when it comes to evaluating changes in health warnings. Unless both survey waves are conducted when the same set of health warnings is on the pack, the baseline measures of processing or knowledge relate to the "old" warnings, whereas any cessation-related activity at follow-up presumably reflects the impact of the "new" warnings.

An alternate strategy that can also be used in cross-sectional studies is to ask smokers to directly report the extent to which warnings have influenced their motivation to quit and smoking behaviour (Borland & Hill, 1997a;

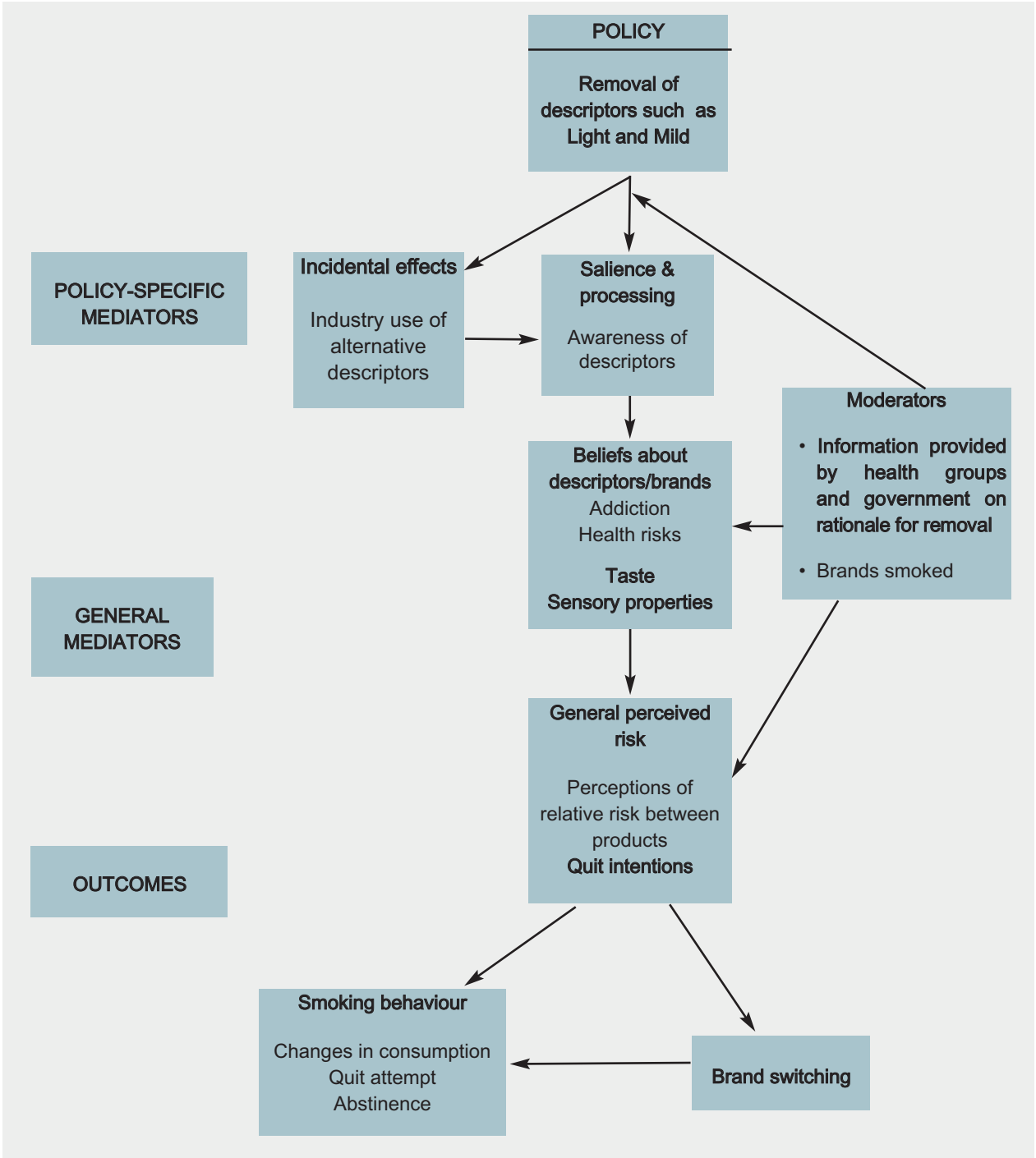


Figure 5.31 Conceptual Framework for the Evaluation of Brand Descriptor Policies

Canadian Cancer Society, 2001; Health Canada, 2005; Koval *et al*., 2005; Willemsen, 2005; O'Hegarty *et al*., 2006). This approach does not have the same validity in terms of measuring actual changes in smoking behaviour, although it can be used to examine changes across labelling policies.

A third alternative is to examine changes in prevalence rates, or population-based cessation activity, before and after the implementation of new warnings. To our knowledge, this approach has been used in only one study to date: Gospodinov & Irvine (2004) reported no discernable changes in prevalence rates, and a reduction of two cigarettes per week among smokers in the months following the implementation of pictorial health warnings in Canada. However, as described earlier in this section, there are serious problems in attributing changes in national level trends to changes in health warnings, or any other individual policy measure. Indeed, as Gospodinov & Irvine note, there were significant changes in price over the same period of time, as well as considerable sub-national tobacco control activity over the same time period.

Yet another approach to measuring the impact of warnings on cessation behaviour has been to look at changes in the use of cessation services as they relate to information on warnings labels. Research conducted in the UK and the Netherlands has examined changes in the usage of national telephone helplines after the contact information was included in package health warnings. Each of these studies reports significant increases in call volumes (Willemsen *et al.*, 2002; Department of Health, 2006).

Finally, several items have been created for use among former-smokers. Typically, these items ask about various reasons for quitting, including whether the health warnings either motivated them to quit or have helped them to remain abstinent (Canadian Cancer Society, 2001; Hammond *et al*., 2004b; O'Hegarty *et al*., 2006; Thompson *et al*., 2006). These measures are, however, subject to recall bias and should be interpreted with particular caution (Table 5.39).

Formative research

Formative research is often conducted to help identify the content and design of new health warning policies. Regulators must decide what health effects to communicate, how many, and how to present this information to smokers on the package. Although population-based surveys may help to guide these decisions, qualitative research is typically undertaken as part of the policy development process.

The most common approach has been to conduct a series of focus groups (i.e. semi-structured interviews conducted within a group setting). Focus groups have two important advantages over population-based surveys: 1) participants can be presented with visual stimuli, including examples of health warnings in a way that is not possible with telephone based surveys; and 2) focus groups are well suited to open-ended questions and allow for more in-depth discussion than structured surveys. In many cases, focus groups are also used as a way to evaluate the effectiveness of health warnings on sub-groups, including younger smokers and those from lower socio-economic groups. The primary disadvantage of focus groups is that the findings can be hard to summarize in a systematic fashion, which complicates comparisons across groups and settings. As a result, conventional validity tests for quantitative data can not be conducted with focus group findings. Nevertheless, qualitative findings help to complement quantitative research in this area, and represent an important step in the development of new labelling policies.

Qualitative research has examined many of the same themes as population-based surveys, and other quantitative methods. These include general knowledge of the warnings, such as the content and location, the emotional impact of warnings, as well as their general salience and noticability (Environics Research Group, 2000; Elliot & Shanahan Research, 2002; CRÉATEC, 2003; BRC Marketing & Social Research, 2004; Health Canada, 2006). In many cases, these studies have presented different health warnings to participants in order to make direct comparisons between labelling policies. These designs have proven particularly effective at comparing the emotional reactions

elicited by picture versus text warnings, for example (Environics Research Group, 2000; Elliot & Shanahan Research, 2002; BRC Marketing & Social Research, 2004). Focus groups have also provided critical information regarding the meaning and comprehension of the information communicated in labelling policies. For example, focus group measures developed, on behalf of Health Canada, have helped to demonstrate that, even though most Canadian smokers are aware of emission information on the side of packages, very few understand the actual meaning of the information (Environics Research Group, 2003). Indeed, judging by the findings of the focus group, most Canadian smokers are misusing the emission information. Thus, carefully constructed focus group measures can provide "deeper," more comprehensive measures of meaning that are difficult to ascertain through structured population-based surveys.

Industry documents

Internal tobacco industry documents represent a potentially rich source of information about the effectiveness of tobacco control policies. There are several informative reviews of industry activities and documents on product labelling, including many related to brand descriptors such as "light" and "mild" (Slade, 1997; Pollay, 2001; Pollay & Dewhirst, 2002; Wakefield *et al.*, 2002; Chapman & Carter, 2003; Alechnowicz & Chapman, 2004). However, to date, no comprehensive review of packaging issues related to labelling policies has been undertaken.

Summary

Few of the measures used to evaluate warning label policies have undergone formal psychometric analyses. Much of the literature in this area has been conducted on behalf of regulators, which may account for the lack of "formal" tests of validation more common to academic research. In addition, different studies have used different measures to assess the same construct. In many cases, measures differ in the wording of questions and in the time references used in measures, such as noticing and awareness. This complicates comparisons across surveys and across labelling policies. However, most measures have high face validity and several have shown good predictive validity for downstream outcomes, including knowledge of health effects and self-reported motivation to quit, and cessation behaviours. In addition, the consistency of the findings across studies and survey modalities suggests that the differences in the measures have only a modest effect on outcomes of interest. Nevertheless, virtually all of the constructs would benefit from further developmental work, including the standardization of the wordings across surveys.

Implications for study design & analysis:

No single study research design is adequate to evaluate the impact of labelling policies. Given the challenges inherent in evaluating national level policies, individual studies are inevitably subject to a range of limitations. However, when taken collectively, the range of designs constitute a persuasive body of evidence demonstrating the effectiveness of comprehensive health warnings. Qualitative methods, including focus groups, are essential for informing the early stages of design and generating new insights into labelling policies. Experimental research is best suited to drawing direct comparisons across warnings and to isolating the effectiveness of individual design and content features. For this reason, experimental research provides the highest level of internal validity. Alternatively, population-based surveys have the highest external validity and may provide the most comprehensive measures of effectiveness given adequate designs. External validity is particularly important in the case of warning labels, which operate over repeated exposures that are tied to smoking behaviour. The pattern of exposure is the defining feature of product warnings and one that is impossible to replicate in a "laboratory" environment. As a result, the central question of whether labelling policies influence beliefs, attitudes, and behavioural change can only be

assessed with population-based surveys. The inferences that can be made from these surveys are considerably enhanced within longitudinal and quasi-experimental designs, as discussed in Section 2.1.

Priorities for future work:

As countries begin to implement restrictions on misleading packaging elements, research must begin to examine elements other than "light" and "mild" brand descriptors. These include other potentially misleading elements, such as the use of colour-coding and package designs that falsely convey differences in strength. To date, very limited work has been conducted outside the tobacco industry on these issues. There is an immediate need to develop measures that can examine these issues within population-based samples, especially within jurisdictions where "light" and "mild" descriptors have already been prohibited.

A second priority for future research is to examine contents and emission information more closely. Up to now, much of the existing research has focussed upon awareness and understanding of ISO tar and nicotine numbers. There is an urgent need for measures to evaluate new approaches to communicating contents and emission information. Population-based studies should be conducted within jurisdictions that have developed novel policies, such as communicating emission information using descriptive, rather than quantitative means. Greater experimental and qualitative work must also be undertaken to explore how smokers interpret and use this information, and to compare different approaches more systematically. These issues are directly relevant to the ongoing debate regarding how to communicate the risks of combustible versus non-combustible tobacco products. Historically, emission information has been used by smokers to evaluate the relative risks of different products. As emission and content labelling policies are developed for the full range of tobacco products, regulators will need to consider the delicate issue of what fundamental message they wish to communicate to smokers. Quantitative emission and content information will inevitably be interpreted as indicators of risks, unlike descriptive information that is uniform across products.

In addition to developing new survey measures, existing measures must be administered more widely, as a greater number of countries prepare to implement the provisions within Article 11 of the FCTC. In particular, few of the measures reviewed in this section have been assessed among smokers in low- and middle-income countries.

Finally, measures should be developed to examine the impact of the cessation information that is included in many labelling policies. Cigarette packages are among the most prominent vehicles for disseminating cessation services and efficacy-related information. These measures may include survey based measures, as well as indicators from other data sources, such as usage rates from telephone quitlines or web-based services.

Recommendations

Comprehensive evaluations of health warning labels should include recommended items from each of the key constructs (see above). Population-based surveys, seeking a more limited evaluation of health warnings, should include proximal measures of noticing, along with intermediate measures of perceived risk or health knowledge. Although measures of general awareness and knowledge of health warnings can be informative, these measures should be used with caution for the purpose of comparing labelling policies.

Evaluations of brand descriptors, and other packaging elements, should represent a priority for tobacco control policy. In addition to examining "light" and "mild" descriptors, research should consider other potentially misleading terms, as well as brand elements such as colour and package design. Unlike health warnings, these policies require the removal of information from the package and present challenges in the wording of survey measures. There is an immediate need to develop measures that can address these issues as more countries implement recommendations under Article 11 to

prohibit misleading package elements.

Policies to communicate emissions and content information via packages, also present unique evaluation challenges. Unlike health warnings, measures of salience and processing for this type of information are of limited value. Rather, evaluations should focus upon the meaning and use of emission and content information. Given the lack of research in this area, and the lack of consensus regarding the best policy approach, there is a particular need for formative research in this area.

Overall, the selection of measures to evaluate tobacco labelling policies will depend upon the method and scope of the evaluation, as well as the specific policy context.