Summary

Background

In the 20th century, cigarette smoking caused an estimated 100 million deaths worldwide (Gajalakshmi *et al.*, 2000). Most of these deaths were in high-resource countries where cigarette smoking first became popular in the 1920s to 1940s. This resulted in an epidemic of smoking-induced cancer, heart disease and chronic obstructive pulmonary disease (COPD) deaths. Cigarette smoking is not only the most prevalent form of tobacco use, it is also particularly harmful, killing one of two long-term users, half of them (one in four users) in middle age. In 2000, smoking was responsible for approximately 4.83 million deaths in people 30 years of age and older, evenly divided between high- and low-resource countries (Ezzati & Lopez, 2003), with lung cancer accounting for 0.52 and 0.33 million deaths, respectively (Ezzati & Lopez, 2004). If current mortality trends continue, it will cause some 10 million deaths each year by 2030, with around 70% in low-resource countries (Peto & Lopez, 2001). If present usage patterns persist, smoking will cause approximately 1 000 000 000 deaths this century, a tenfold increase over the previous century (Peto & Lopez, 2001). Most of these expected deaths could be averted if we rapidly institute effective pro-grammes to both discourage tobacco use and to assist those addicted to tobacco to quit (IARC, 2007a).

Tobacco is a plant containing the psychoactive and addictive drug nicotine. Although nicotine is the main psychoactive ingredient of tobacco and the source of its addictiveness, it is otherwise a minor contributor to the harm (Benowitz, 1998). Most of the harm is due to other constituents in tobacco, particularly in tobacco smoke (IARC, 2004). The harms from tobacco mainly stem from long-term use, which the addictive nature of the product promotes.

Across its long history, tobacco has been processed and consumed in a wide variety of ways. The two main forms of use are smoking combusted tobacco, and taking unburned tobacco into the mouth or the nose (smokeless use). Over the 20th century, the use of cigarettes, primarily factory-made cigarettes, dominated tobacco markets in nearly all countries. Cigarettes have also been the focus of most tobacco research. The use of other smoked tobacco products is now of only minor importance, except in some areas, particularly the Indian sub-continent, where the use of bidis prevails. All forms of smoked tobacco are very harmful to health (IARC, 2004), and attempts to create less-toxic versions of these products have generally failed, largely because they have been unacceptable to consumers. Smokeless tobacco, which is generally less harmful than smoked tobacco because it does not involve inhaling smoke, but still carcinogenic to the oral cavity and pancreas (IARC, 2007b), is not used in many parts of the world, but it is common in some areas and its use is significant and increasing in some countries (e.g. Sweden; Foulds *et al.*, 2003). With some forms of smokeless tobacco there has been success in reducing toxins while maintaining consumer acceptability (Broadstock, 2007). Non-cigarette tobacco use is under-researched in comparison to cigarette use.

In recognition of the threat that tobacco use poses to global public health, in May 2003, the member countries of the WHO adopted the Framework Convention on Tobacco Control (WHO FCTC), the first international treaty devoted to improving public health by restraining tobacco promotion and use (WHO, 2003).

Scientific evidence plays a central role in the WHO FCTC. Its Foreword describes the WHO FCTC as "an evidence-based treaty that reaffirms the right of all people to the highest standard of health" (WHO, 2003). The preamble to the FCTC states that adopting nations are "determined to promote measures of tobacco control based on current and relevant scientific,

technical, and economic considerations" (WHO, 2003). To achieve its objective, the WHO FCTC calls for a comprehensive range of policies, defined for the purposes of this Handbook as the enabling mechanisms that allow particular rules, regulations and programmes to operate (in other words, frameworks that allow instruments to be implemented). The key articles of the Convention relevant to this Handbook are:

Article 6
Price and tax measures to reduce the demand for tobacco

Article 8
Protection from exposure to tobacco smoke

Article 9
Regulation of the contents of tobacco products

Article 10
Regulation of tobacco product disclosures

Article 11
Packaging and labelling of tobacco products

Article 12
Education, communication, training and public awareness

Article 13
Tobacco advertising, promotion and sponsorship

Article 14
Demand reduction measures concerning tobacco dependence and cessation

Article 15
Illicit trade in tobacco products

Article 16
Sales to and by minors

Article 17
Provision of support for economically viable alternative activities

Article 20
Research, surveillance and exchange of information

Article 22
Cooperation in the scientific, technical, and legal fields and provision of related expertise

The WHO FCTC is a seminal event in global health. Scientific evidence has demonstrated the enormous health harms of tobacco use. Scientific evidence as to the effectiveness of potential interventions formed the basis for the selection of the policies that are included in the WHO FCTC. However, whether the WHO FCTC is to fulfill its objective of reducing the devastation of the tobacco epidemic will depend on how effectively countries formulate and implement these policies. Moreover, history has shown us that the tobacco industry will adapt and work to circumvent even the strongest policies, so governments will also need to be ready to evolve and change their policies in order to ensure they achieve their goals. Good public health practice demands ongoing evaluation research as critical to informing the implementation and

dissemination of established policy instruments as well as to aid in the subsequent evolution of new policy-related interventions.

Overview

This Handbook is concerned with the articulation of a framework and methods for conducting tobacco control policy evaluation, and not with an evaluation of a body of research in itself. It also offers terminology to judge the quality of the evidence considered in such evaluations and to be applied by IARC in the future evaluation of specific tobacco control policy interventions. As a result, the WG's advice to the potential readers of the Handbook is largely about how to evaluate policy interventions in ways that we believe will best advance tobacco control. In addition to this advice to researchers and evaluators, a small number of recommendations directed at other audiences are made.

The goals of this Handbook are to move the field by:

a) developing a common framework and language for tobacco control policy evaluation;
b) reviewing the strengths of possible research designs;
c) using theory to derive core constructs to measure when doing evaluations of key tobacco control policies;
d) identifying measures of constructs, and
e) providing an assessment of the scope and quality of existing

data sources. Four broad questions guided the review of the scientific literature on the methods and measures of tobacco policy evaluation:

1. How do we determine the effects of a policy?

   What are the key features of the policy as implemented?

   Is there a common conceptual framework that can be applied to understand how policies work?

   How might different design features be used to reduce threats to internal validity?

2. What are the core constructs for understanding how and why a given policy works?

   Which of these are parts of general pathways, and which are specific to particular policies?

   What is the quality of the measures used to assess core constructs?

   Do these measures, as well as the constructs they presumably reflect, translate into different cultures and contexts?

3. What are potential moderator variables to consider when evaluating a given policy?

   What is the quality of the measures used to assess potential moderator variables?

4. What data sources exist that might be useful for evaluation?

   How useful are these data sources for evaluation (i.e. completeness and quality)?

The WG acknowledged that in attempting to answer these questions, explicit considerations must be given to equity issues both within and between countries. This involves always asking the question: "What is needed to optimise the intervention for disadvantaged groups?" This may range from making sure a programme is available in disadvantaged areas, to ensuring that the wording and tone of communications is acceptable and comprehensible.

The Handbook outlines a framework that interested organisations, including governments, can utilise to measure the effectiveness of interventions aimed at implementing tobacco control policies that are currently being and will be adopted in the next several years in adherence to the WHO FCTC. It describes major steps we made to articulate a new and coherent framework for thinking about tobacco control interventions.

The WG came from diverse disciplines, with different theoretical traditions and methodological approaches. This necessitated ongoing work to standardise language. We realised that some terminology was designed for thinking about the problem from a different perspective to the one necessary for understanding the complexity of population health areas like tobacco control. There is a need for ongoing work to rethink our terminology to better fit a population health framework.

The Handbook is intended to be a resource for researchers interested in evaluating tobacco control policies, and others interested in evaluating interventions beyond merely auditing implementation. It should also be useful for policy and programme developers as it spells out the theoretical frameworks upon which the interventions are based, and provides explicit models of how they exert their effects.

Steps towards a framework for evaluation

The WG began by considering what outcomes to focus on. It concluded, insofar as the interventions under consideration related to tobacco use and not to the harmfulness of each unit of the product, that the focus should be on tobacco use behaviours as the main outcomes of interest. This meant that, for the most part, the WG did not consider disease or mortality outcomes.

The WG concluded that there is currently no coherent framework for thinking about the evaluation of tobacco control policies in the policy literature. The frameworks borrowed from other areas such as clinical medicine are not adequate to the needs of the policy field. Randomised clinical trials are neither necessary nor often practical to generate evidence of the effectiveness of tobacco control policies.

The WG concluded that policy evaluation should be conceptualised in a manner analogous to how epidemiologists approach the task of inferring conclusions about the causes of disease (US Department of Health, Education and Welfare, 1964; Hill, 1965). This

is a framework that encourages researchers to triangulate all the available evidence to help rule out alternative explanations of observed effects, rather than focus on attempting to draw conclusions only from individual studies or from meta-analyses of studies using the same study design.

In the same way that evidence-based medicine has been built from rigorous evaluation of treatment options, evidence-based public health must begin with building a database from rigorous evaluation of public health policies. Evaluation of the effectiveness of tobacco control policies at the population level has been limited by inadequate data sources, problems in measurement and poorly conceptualised evaluation designs. It has also been limited by a failure to look for and maximise the value of studies with individually limited designs by systematically reviewing the findings from the corpus of such studies to determine what they collectively add to knowledge. In isolation or even combined in meta-analyses of similar studies, they may have little to tell us, but when they are combined in ways that take account of different threats to the validity of attributing causality by study type, they can sometimes be used to make strong inferences about causality as well as potentially increasing our understanding of the conditions under which the interventions are most effective. The benefits of such an approach are not just with regard to increasing our understanding of the effects of the intervention, but it also
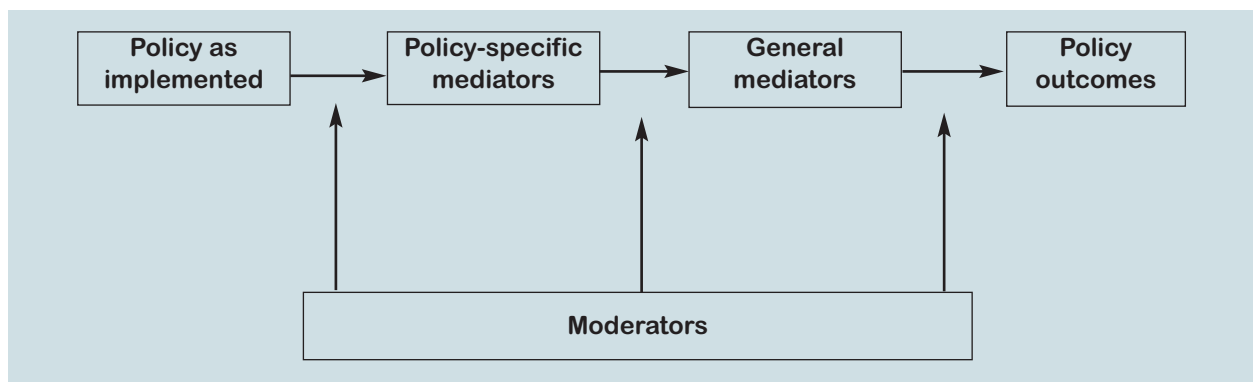
improves our ability to understand individual cases. Explicit comparison with the corpus of existing knowledge allows individual evaluators to say more about the programmes they evaluate than the designs they have adopted would allow them to do if they treated their evaluations in isolation of the accumulated knowledge.

The question one usually asks about policy interventions is: "Under what conditions can the desired effects be optimised?", not whether the intervention can work. Translated to the individual case, the question becomes: "Is the intervention working here as well as it should?" To answer that question one must be concerned about the form of the intervention, the ways it is delivered (quality of implementation), and various characteristics of the populations it is addressing. This is a framework that sees evaluation as part of a process of continual improvement. It is also about determining the relative contribution of each intervention to the overall goal, and how this might be moderated by characteristics of the broader environment. It involves paying more attention to the articulation of theoretical mechanisms, and having study designs that facilitate the elaboration of causal mechanisms.

Good evaluation starts with an analysis of the problem. Thus, the need to build an understanding of the factors that are affecting or can affect tobacco use and how use relates to the harms. Mechanisms by which tobacco control interventions can act to reduce harm must also be considered. The

WG identified four aspects that need to be considered in evaluating interventions designed to reduce the harms. First, one must consider whether the goal of the intervention is to change tobacco use, tobacco harmfulness, or both of these. Second, a theoretical model or set of models describing how the interventions are expected to achieve their intended effects must be developed. Third, possible incidental effects of a policy that may occur must be considered. Fourth, any change in the environment that could modify the impact of the intervention (particularly counter-actions of the tobacco industry) must be monitored, and evaluated where necessary.

The first three steps in determining how policies may achieve their effects require specification of a theory of how the policy is expected to work. As Kurt Lewin noted years ago (1935), "there is nothing as practical as a good theory." The WG concluded that researchers should consider the adoption of a common framework to help identify relevant theories and thus guide the selection of core constructs useful for evaluating how and under what conditions tobacco control policies work. The issues that are likely to be relevant are to be considered well in advance. A general framework for assessing how an intervention might work is illustrated in Figure 6.1. At the first level it specifies two levels of mediating variables between a policy intervention and the outcomes: those specific to the policy, and those variables that are part of more

| Policy as implemented | → | Policy-specific mediators | → | General mediators | → | Policy outcomes |

| Moderators |

Figure 6.1 A generalised model of mediation making allowance for moderator effects

general pathways. It also accepts that various other factors (moderators) might affect the size of the effect.

There are only two main types of causal chain one needs to consider: the pathway from policies to tobacco use, and the pathway from tobacco products to levels of exposure to toxic substances and to the harms that result. Consideration of pathways may lead to the subdivision of a policy area into classes of interventions that share common pathways.

Understanding the mechanisms by which interventions have their effects is important because: 1) it can provide strong evidence of the causal impact of a particular policy, especially when attempting to differentiate the effects of a specific intervention from other possible causes, including other tobacco policies; 2) it can be used to diagnose the problem in cases where intended effects did not occur, by identifying where in the causal pathway things went wrong; 3) it can help us understand why a

policy does not have the intended effect for some groups, but does for others (i.e. clarify why moderation occurs); and 4) in specifying how a policy works, it may help identify alternative ways of achieving the desired effects. These understandings can facilitate the development of new, and hopefully improved, ways of targeting key pathways of influence, or of tailoring interventions to better reach more resistant or needy groups.

The model outlines the primary constructs involved in helping to explain the relationship between tobacco control policies and their effects on tobacco use behaviours. In a limited number of cases, primarily in some aspects of product regulation, there is an alternative main path to outcomes—through reduced delivery of toxic chemicals. This is spelled out most clearly in the section on product regulation.

It is particularly important to go beyond the specific intent of some policies to explore their more distal ramifications. For example, the

goal of information and product labelling policies is improved dissemination of knowledge to the potential user of the product. However, it is of interest to see whether and how these policies actually translate into changes in tobacco use behaviours. It is also important to consider effects along different pathways to the intended means of action, as these might be important for analysis of society-wide effects; e.g. the generally neutral or positive effects on business of smoke-free policies.

Finally, there needs to be consideration of unexpected effects on other determinants of tobacco use. This consideration is more important in tobacco control than in most other areas of health because such effects may be deliberately influenced by the tobacco industry (Cummings *et al.,* 2002b). Hence, surveillance of tobacco industry practices is required. The approach taken can be facilitated by a theoretical understanding of the industry's profit motive and marketing

practices, as this can guide the selection of data that are most relevant in surveillance for counter-active effects.

The conceptual framework for behaviour change assumes that each policy directed at changing tobacco use ultimately has its influence on those behaviours through a specific causal chain of psychological events. Policy-specific mediators involve such things as awareness, policy-specific knowledge and reactions to specific elements of the intervention. For example, new graphic warning labels should increase salience and visibility of warnings, and perhaps foregoing of occasional cigarettes. The second set of general mediators are constructs taken from behavioural science that we know mediate effects of behaviour; that is, they are means by which changes in tobacco use may occur. They include attitudes, normative beliefs and intentions. Moderators—those things that change the magnitude of the effects of an intervention without necessarily being changed by the intervention—include socio-demographic factors (e.g. age, gender, socio-economic status, cultural background) and psychological factors that are either assumed to be stable or which the intervention is not designed to change (e.g. level of dependence). This framework provides a general guide for thinking about policies and their effects on a broad array of important psychosocial and behavioural variables, and for testing

how differences in policy implementation relate to effectiveness.

The model for the effects of changes in tobacco products to health effects can similarly be articulated, although here the distinction may be more between constructs that are measured in the environment (e.g. physical characteristics of cigarettes) and those within the individual (e.g. exposures, health harms), and the challenges of demonstrating links between the two—for example, the failure of current measures of cigarette yield to relate to measures of exposure to those chemicals in smokers.

The WG set the task of using diagrams, or logic models, to spell out the main factors to consider for each policy area and how they interrelate when considering all these policies simultaneously; to see if this approach would help elucidate common constructs and measures that might apply across different policy domains. The logic models allowed the WG to readily compare the similarities and differences in the constructs and measures across policy domains, and of the differences of policy type within a broad policy domain. The models were deliberately kept simple in an effort to focus attention on key constructs.

Finally, a major challenge is in the identification and validation of appropriate measures. Measurement validity is a particular issue, with measures of constructs varying in their validity dependent on the purpose they are being used for. This is sometimes because

measures of known bias are used for measuring constructs because no better measures exist, but the differential effects of that bias in different contexts are overlooked.

The general theoretical framework presented here should be applicable across socio-cultural contexts. Clarification of policy intervention effects and the moderation of these effects will often involve comparative research. However, the specific theoretical model, its associated constructs and the measurement of these constructs may differ in important ways across national, cultural, linguistic and social groups. Where this happens, caution must be exercised in making comparisons between such groups.

Section Summaries

General methods and common measures

The Handbook first discusses features of research design for evaluation studies and how those features can form the basis for stronger conclusions about the impact of policies. Other aspects discussed and deliberated on include measurement issues in the design and analysis of cross-cultural comparative research, as well as some of the methods currently recommended for attempting to resolve these issues.

The Importance of Design in the Evaluation of Tobacco Control Policies

Evaluating the outcomes of population-level tobacco control policy involves three interrelated questions:
(1) Does the policy have an impact? (causality); if so,
(2) Under what conditions? (moderation); and
(3) How (mediation)?

The choice of design elements will depend on which questions are considered to be a part of the evaluation effort.

It is important to ensure that the appropriate concepts are chosen and that for each, measures are identified that are suitable to answer the evaluation question.

This section describes key design elements of outcome evaluation studies and how each contributes to reducing or eliminating threats to the internal validity of a study. Internal validity determines the extent to which the results of the study can lead to a causal conclusion.

Evaluation efforts should be informed by knowledge of the nature of the policy being evaluated, and the goals of the evaluation study should be clearly stated. Evaluation planning should be guided by understanding what threats to internal validity may be present in the study of a given policy situation, and then adding design elements and other measures to reduce or eliminate those threats.

Knowledge of the mediational pathways that are theorised to explain how policy affects behaviour and environment (or environmental risk) should lead to an appropriate study design, the inclusion of appropriate constructs and measures, and the selection of analytic tools that are well-suited to estimating the causal impact of policies by providing an explanatory pathway and helping to eliminate alternative explanations. Logic models describe these pathways and help identify constructs to measure. Suggestions on specific measures for many of these constructs are provided in other sections of this Handbook.

An outcome evaluation study must, at a minimum, include one post-policy measurement. In general, the addition of one pre-policy measurement (even cross-sectional) using the same measures and sampling frame is a more powerful evaluation strategy for assessing change due to a policy. The inclusion of a single, non-random control from another population is considered less desirable. Additional post-policy measurements are useful to track the effects of a policy over time. The utility of longitudinal designs is strengthened if there are multiple data collections before and/or after policy implementation, as this allows more precise specification of effects—for example, taking into account temporal trends that were occurring before the implementation of the policy. The role of time series analysis on aggregate sales/consumption data to demonstrating the effects of price on consumption is a good example of the power of multiple measurements.

Both repeated cross-sectional and longitudinal (cohort) designs are useful for assessing the impact of a given policy. The use of cohort designs provides additional capability for tracking the impact of policies within individuals, allowing stronger tests of mediational pathways.

Addition of samples from other populations to either or both intervention and control arms also adds strength to the evaluation design, as does having varying levels of intensity of the intervention.

Similarly, parallel assessment of alternative explanations for observed changes in outcomes (e.g. possibly being due to other policies or industry counter-actions) adds strength over assessing these effects in separate studies.

The existence of studies with complementary strengths and weaknesses is particularly useful in triangulating the results of a corpus of evaluation studies to see if a consistent pattern emerges.

The use of probability sampling in an evaluation study increases its external validity—the extent to which the findings of a policy evaluation study can be generalised to making conclusions about the impact of the policy on the larger population.

At a broader level, the design of an evaluation study should be guided by knowledge of how prior evaluation studies in the same policy domain have been conducted. An analysis of the similarity or differences in policy impact across similar studies can yield powerful conclusions about the overall impact of a policy.

Developing and assessing comparable questions in cross-cultural survey research on tobacco

Evaluation of tobacco control policies and other population-level interventions often involves data collection efforts across diverse national, cultural, linguistic and social groups. Comparison across such groups is often necessary to clarify policy effects, how these effects happen and how effects might differ across populations. The literature discussed in this section suggests that these comparative studies should consider measurement equivalence issues in the following ways:

Research teams should include collaborators from the socio-cultural groups in which the study is being conducted in order to help anticipate issues regarding the comparability of the theoretical framework, constructs and the measurement of these constructs across groups. When research involves participants from distinct language groups, it is recommended that at least one, and preferably more, team members are fluent in the source language and the target language in which the survey will be administered.

Whenever possible, it is recommended to use measures that have been appropriately validated for the populations in which the questionnaire will be administered. Even when a measure has been validated within one population group, its validity may not extend to other groups, and additional steps may

be necessary to increase validity and improve the value of comparisons across groups.

Translation of questionnaire items from one language to another should involve experienced translators. Review and adjudication of multiple, independent translations of the same items is currently considered the gold standard. If only one person translates the questionnaire, translation review should involve a group of bilingual people who are knowledgeable about questionnaire design principles and key study concepts. Translation assessment should not merely consist of back-translation.

Researchers should carefully select and translate items with the goal of achieving equivalence of construct meaning across study populations. In some cases, literal translation of a questionnaire item across linguistic variants of the survey will not adequately capture the construct of interest, and more flexible translation and adaptation of the question will be necessary.

All surveys, not just those that are translated, should be pre-tested to assess comprehension issues among the populations in which the survey will be administered. Ideally, pre-testing would involve cognitive interviewing before a survey is fielded. Cognitive interviewing or other pre-testing methods may also be used post-hoc to increase the validity of comparisons or to determine whether inconsistent results may be due to differential question comprehension.

Researchers should consider and seek solutions to minimise the

ways in which culturally moderated response factors (e.g. social desirability, acquiescence, extreme responding) may influence responses.

Researchers should document decisions related to measurement development and item wording, especially where conceptual equivalence is suspect, translation is difficult, or where cognitive interviewing or other pre-testing methods reveals systematic differences in meaning. Researchers should also document issues around survey administration.

Outcomes and major determinants

Next, the Handbook presents constructs that are likely to be used across a range of policy evaluations, factors that can influence the validity of self-report tobacco use behaviours, factors that can influence comparability across surveys, and measures to assess use, providing examples from cross-national surveillance and evaluation systems as well as national sources. A core set of general mediator and moderator variables that may be relevant to consider in evaluations of tobacco control programmes and policies, with a brief description and assessment of some standard measures for assessing these constructs, are discussed. Self-report measures of nicotine/tobacco dependence in adults, concentrating on measures that are potentially appropriate for population-based/epidemiologic research, are reviewed as well.

Measuring Tobacco Use Behaviours

The Handbook describes the key concepts within the natural history of tobacco use, providing a conceptual model to guide measurement of key constructs. Current tobacco use is the most important construct because of its importance as an outcome in policy evaluation studies. Studies that have examined the validity of self-report measures of current use generally find these measures to be valid, although there exist some conditions under which the validity may be reduced.

It is important to measure the type of tobacco used, particularly in those countries in which a variety of types exist. The variety of forms available, the possibility of switching, or multiple concurrent use may influence the probability of quitting and disease risk.

Detailed measurement of information about tobacco product packaging is important in order to determine the variant of product type used, movement between price sectors and, potentially, to assess the use of tobacco from illicit sources.

Other important constructs in the measurement of tobacco use behaviour include early use, frequency and quantity of current use, quit attempts and duration of abstinence among former smokers.

Consumers of survey data in which tobacco use measures are included should be aware of factors that can influence population estimates of tobacco use, and take those into consideration when comparing estimates from surveys conducted within and across countries.

Measuring the Psychosocial Determinants of Tobacco Use and Dependence

The WG describes mediators and moderators theorised to be important in understanding how policies and interventions affect tobacco use behaviours, and under what circumstances they have an impact. A core set of measures likely to be important has been identified. Researchers should select from this list and, when appropriate, supplement it with other relevant measures, depending on the specific context and aims of each study. There are validated measures of many of the reviewed constructs, and researchers should whenever possible use these measures rather than developing their own ad hoc measures. Investigators should report the psychometric properties of their measurement instruments, reporting at least test-retest reliability, convergent validity and/or predictive validity. Psychological measures are particularly sensitive to wording and to cultural context, so we recommend that the methods for translations and cultural adaptations described elsewhere in the Handbook be utilised in populations where these measures have not been previously validated.

Measurement of nicotine dependence

Nicotine dependence is an important construct to assess as a moderator for the effects of tobacco control programmes and policies. The WG reviewed the evidence on the validity of various proposed measures of cigarette and smokeless-tobacco-induced nicotine dependence. For cigarette smoking, the 2-item Heaviness of Smoking Index is recommended for use in population-level studies. If only a single item measure is possible we would recommend the use of "time to first cigarette in the morning" as the item. For smokeless tobacco, the Fagerstrom Test for Nicotine Dependence-Smokeless Tobacco (FTND-ST) appears to be a useful measure of nicotine dependence.

Existing data sources

The Handbook then describes sources of details about tobacco control policies, sources of information about tobacco production and trade and repo sitories of youth and adult surveillance surveys. These sources of information are particularly important for making comparisons between countries, and in some cases can be used to demonstrate policy impacts, although not the mechanisms by which they occur.

Data sources for monitoring tobacco control policies

The Handbook describes the new WHO Global Tobacco Control Report (GTCR), a repository of good-quality information on a wide range of tobacco control policies for the large majority of countries. The

GTCR contains copies of most of the legislation and regulations, some measures of scope and/or level of policy enactment, and an indicator of cases where national level policies may mask a diversity of sub-national policies. It is designed to be updated annually.

All policy researchers studying policy differences between countries should use it, and indeed it may be the easiest way to get this information for some individual countries.

The GTCR is limited in what it can provide on extent of implementation and/or enforcement. Its main limitation is that it does not contain information about sub-national policies, as information of this sort is only available for the limited number of countries that collect it.

Data sources on tobacco production, trade and sales

National data on the production, trade (export and import) and sales of tobacco products are most often available publicly at little to no cost and have been underutilised in evaluations of tobacco control programmes and policies. These data 1) can provide important insights into the relevant players and sectors in the national and regional political economy of tobacco control, 2) can be used to construct measures of historical trends in tobacco consumption and 3) provide estimates of the magnitude of the smuggling market. Thus, these data are important information sources for evaluation of tobacco control policies.

National data are typically available from sources such as government statistics agencies and ministries of trade and industry. The United Nations Statistical Division (UNSD) consolidates this information based on reports from countries. These reports are generally accurate, but primary sources should be used to confirm the data and to obtain other information such as data on sales and other tobacco products.

Data sources for monitoring global trends in tobacco use behaviours

The youth surveillance systems described in this section include The European School Survey Project on Alcohol and Other Drugs (ESPAD), the Global School-Based Student Health Survey (GSHS), the Global Youth Tobacco Survey (GYTS) and the Health Behaviour in School-Aged Children Survey (HBSC). The adult surveillance systems described include the Global Adult Tobacco Survey (GATS), the International Tobacco Control Policy Evaluation Survey (ITC) and the STEPwise Approach to Chronic Disease Factor Surveillance (STEPS).

To evaluate articles of the WHO FCTC among youth, GYTS is the only source of international data available that includes the following indicators: exposure to secondhand smoke, exposure to pro- and anti-tobacco media and advertising, cessation, minors' access and school curriculum.

To evaluate articles of the WHO FCTC among adults, GATS and ITC have the most com-

prehensive set of indicators, including: exposure to second-hand smoke, economics (price and taxation), cessation, product labelling, and exposure to pro- and anti-tobacco media and advertising. Where possible, longitudinal studies such as ITC should be used for evaluating policies and programmes because of the opportunity to examine and adjust for individual level predictors of tobacco use behaviours.

GYTS was developed, and GATS is being developed, for countries that did not have existing surveillance systems for the collection of information on tobacco use and its determinants.

Strategies for evaluating specific policy domains

The final section of the Handbook covers all major domains of tobacco control policies except for prevention policies and illicit trade. Here it is illustrated ways in which logic models can be used to highlight the different foci of policies. In particular, analysis of policy areas directed at controlling tobacco marketing (including some forms of product regulation) have identified the importance of monitoring tobacco industry innovations designed to mitigate the policy effects, while those less targeted at the industry have not done so.

Measures to Assess the Effectiveness of Tobacco Taxation

Article 6 of the WHO FCTC calls for ratifying nations to reduce the demand for tobacco products through taxation policies and other product price-related policies. This section focused on the measures needed for evaluating the impact of tobacco taxation, a highly effective tool for reducing tobacco use. The impact of tobacco taxes on tobacco use behaviours is mediated by tobacco product prices, tobacco company price-related marketing efforts, tobacco users' purchase behaviour, tax avoidance and smuggling.

Measuring tobacco product taxes is straightforward, with information on the level and structure of these taxes readily available from the Ministry of Finance and other sources (e.g. the International Monetary Fund, the WHO's GTCR). In some countries, it will also be important to measure sub-national taxes. Three methods for measuring tobacco product prices are discussed in this section: technology-based, observational and survey-based. These methods have differing strengths and weaknesses, and their costs will vary considerably.

To the extent that a national measure of price is of the most interest and a regularly repeated population survey of tobacco use is in place, including questions on price in such a survey would be most efficient. Measuring tobacco product purchase behaviour can be easily done through the addition of a limited set of questions to this survey. Developing accurate measures of tax avoidance and tobacco product smuggling is more challenging, and the validity of these measures is unclear and needs further research. Some of the questions on purchase behaviour in population surveys can be used to provide a range for the extent of tax avoidance. Multiple methods, most of which have not been widely applied and which need further research, can be used to assess the extent of tobacco product smuggling.

Measures to assess the effectiveness of smoke-free policies

Article 8 of the WHO FCTC calls for ratifying nations to adopt smoke-free policies for public indoor locations and workplaces. Evaluating the effects of public smoke-free policies is critical to understanding how these polices are implemented, whether they reduce exposure to tobacco smoke, and how they can be improved. The core constructs identified for evaluating smoke-free policies include compliance with the policy and exposure to tobacco smoke. Based on our review of the available research literature, we conclude that population surveys can generally be relied upon to provide valid measures of compliance with a public smoke-free policy and exposure to tobacco smoke. These self-report measures have been validated by ambient air monitoring and biomarkers of exposure to tobacco smoke. The review here also suggests that it may be important for evaluators to consider measuring key incidental effects of public smoke-free policies such as the impact on the behaviour of smokers, possible changes in smoking behaviour in the home and a variety of potential economic effects.

Measures to assess the effectiveness of tobacco product regulation

Articles 9 and 10 of the WHO FCTC call for ratifying nations to adopt policies for the regulation and disclosure of tobacco product contents and emissions. This section focuses on a review of the methods and measures for evaluating policies that are intended to regulate tobacco products. There are currently five main types:
1) regulations that require disclosure of product information;
2) regulations intended to reduce product toxicity and harm;
3) regulations intended to reduce the addictiveness and/or attractiveness of tobacco products;
4) regulations intended to prevent cigarette-caused fires; and
5) bans (or removal of bans) on product categories.

The selection of specific constructs and methods for evaluation will vary depending on the goals of the specific policy. However, as a general framework the impact of tobacco product

regulations on intended health outcomes will likely be moderated by changes in product design and performance, product marketing, product-related beliefs and attitudes, and tobacco use behaviour, which in turn are expected to influence exposures to tobacco constituents and emissions. Thus, evaluations should not be limited to assessing compliance within the intended effects of a regulation, but should also consider unintended effects or responses, such as tobacco industry innovation, that may interfere with the impact of the regulation.

There is a need for a centralised database that would at a minimum characterise different product regulations so that the effects of different policies can be compared. Additionally, as a condition permitting tobacco product sales, governments should require (if they do not currently do so) tobacco product manufacturers to regularly disclose information about their products at the finest level of brand subcategory, including sales and marketing data, product content and design features. This is needed to inform the development, implementation and evaluation of effective regulations. Additionally, ongoing surveillance is required to assess the impact of tobacco product regulation on the tobacco product market and on the population, as well as to detect industry responses and other unanticipated consequences of regulation. The challenges of measurement associated with evaluating the

effects of tobacco product regulations should not be underestimated. For example, many governments have enacted maximum smoke emissions standards (i.e. tar, nicotine and carbon monoxide) based on standardised machine testing protocols for the purpose of reducing exposure to the constituents in tobacco products and resultant harm. However, based on the evidence reviewed in this Handbook, we recommend against using yields from standard machine testing protocols such as the ISO cigarette testing method (ISO Standard 3308, 2000) to assess or predict human exposure. Emission yields derived from these protocols are not valid measures of actual human exposure. In order to evaluate the effectiveness of product regulations aimed at reducing harm, measures of human use and exposure are essential. There is an urgent need to identify valid methods and measures for assessing human exposure and harm that have practical utility for evaluating tobacco product regulations.

Measures to assess the effectiveness of restrictions on tobacco marketing communications

Article 13 of the WHO FCTC encourages ratifying nations to adopt comprehensive tobacco marketing restrictions to the extent constitutionally possible. This section identifies the key issues and constructs for evaluating restrictions on tobacco marketing. Tobacco marketing includes all

the communication efforts tobacco corporations use to encourage consumption of their products, including mass media advertising, sponsorship of sporting and cultural events, point of sale promotion, merchandising and give-aways, and public relations.

A core distinction to consider is between evaluation of the pathway of intended effects, and the need to monitor, and evaluate where necessary, evidence of tobacco industry activity that might reduce the impact of the policy.

Various methods can be used to measure the effects and effectiveness of restrictions on tobacco marketing, some borrowed from strategies to assess the impact of marketing. The main approaches include using consumer surveys to examine the target market's response to bans and restrictions and, if it can be obtained, use of disaggregated tobacco company marketing expenditure data to model changes in tobacco use. Given different limitations, the WG recommends a mix of these approaches, along with others where possible. However, there is a critical need to develop methods and valid measures for estimating the effects of marketing bans and restrictions at the level of the consumer.

Additional key challenges in evaluating the effects of marketing bans and restrictions include the extended time required for past marketing campaigns to dissipate from people's awareness, and the persistence of effects from recent campaigns. Innovative and increasingly subtle tobacco industry

marketing strategies create an urgent need for ongoing monitoring of industry behaviour.

Measures to assess effectiveness of product labelling

The WHO FCTC proposes tobacco product labelling regulations in 3 main areas: 1) health warnings, 2) misleading brand descriptors, such as "light" and "mild", and 3) information on the constituents and emissions of tobacco products (Article 11). The Handbook identifies core constructs for evaluating labelling policies including: proximal outcomes such as awareness, processing and knowledge of health warnings; intermediate outcomes such as health knowledge, perceived risk, affective reactions, avoidance, brand appeal and cessation knowledge; and distal outcomes such as motivation to quit, changes in consumption patterns and quitting behaviours. Few of the measures for each of these constructs have undergone formal validation testing, although several of the measures described have shown utility for evaluating the impact of changes in product labelling.

The selection of specific measures to evaluate tobacco labelling policies will depend upon the policy chosen for evaluation. Evaluations of health warning labels should include proximal measures of noticing, along with intermediate measures of perceived risk or health knowledge. Evaluations of brand descriptors and other packaging elements should be a priority for tobacco control research. Unlike health warnings, these policies require the removal of information from the package and present challenges in the wording of survey measures. Evaluation of policies intended to communicate emissions and content information via packages should focus upon understanding and use of this information rather than knowledge or awareness.

Measuring the impact of anti-tobacco public communication campaigns

The WHO FCTC Article 12 requires ratifying countries to "promote and strengthen public awareness of tobacco control issues, using all available communication tools, as appropriate." Such campaigns seek to increase awareness and knowledge of tobacco-related issues, with the goal of promoting individual behaviour change and support for and progress toward policy and social change. The Handbook provides a framework for evaluating multi-component public communication campaigns in order to design effective campaigns, identify and correct problems of campaigns that are in progress, and to document the campaign's impact. Core methods include testing campaign messages during the design phase, monitoring the reach of the campaign during implementation, and assessing core constructs, including awareness, knowledge, attitudes and beliefs, support for policies and tobacco-related behaviour change. The measures described here, like the campaigns themselves, need to be customised to the specific content, purpose and message of the communication effort being implemented.

Regardless of the results of the public communication campaign (and particularly if it failed to show results), evaluations should be made publicly available. A system to collect and document campaign results would enhance our understanding both of how public communication campaigns work and how to make them better.

Measures to assess the effectiveness of tobacco cessation interventions

Article 14 of the WHO FCTC obligates ratifying nations to adopt policies that promote access to evidence-based tobacco cessation interventions. Such interventions range from less intensive efforts such as brief opportunistic advice by health care professionals to more intensive efforts delivered to tobacco users either individually or in groups by trained health professionals. Core constructs for evaluating access to tobacco cessation interventions include: proximal variables such as awareness of cessation interventions, intermediate variables including specific beliefs and attitudes about different cessation interventions, and distal variables reflecting the utilisation of different cessation interventions.

The effects of policies facilitating access to tobacco cessation interventions can be assessed through self-report using standardised surveys of current and former tobacco users and also by review of records that document trends in the utilisation of tobacco cessation interventions (e.g. calls to a helpline, sales of stop-smoking medications). Measures described here are useful exemplars of how to assess utilisation of cessation services. Evaluations of the effects of policies to promote access to cessation interventions should preferably employ a longitudinal design to assess the relationship between the utilisation of cessation treatments by current and former tobacco users and tobacco use behaviours.