

CHALLENGES AND APPROACHES TO EVALUATING COMPREHENSIVE COMPLEX TOBACCO CONTROL STRATEGIES

Robert Schwartz
University of Toronto, Dalla Lana School of Public Health,
Ontario Tobacco Research Unit
Toronto, Ontario

Gillian Pais
Former research assistant, Ontario Tobacco Research Unit
Toronto, Ontario

Abstract: Challenges to evaluating comprehensive complex strategies revolve around addressing comprehensiveness, attribution, and complexity. The latter requires attention to synergies amongst interventions, feedback loops, and other forms of nonlinearity. The article reviews and assesses how well six approaches to evaluating strategies succeed in dealing with these challenges, including one developed in light of the initial review. None of the approaches offer ideal solutions to the challenges of complexity. The quantified logic model approach suggests the need to simplify and refrain from trying to assess all causal chains in complex strategies. Intervention path contribution analysis, an approach under development, explores the possibilities of using contribution analysis to validate evaluative propositions developed from literature, program theory, and incomplete evaluative information. Greater understanding of synergies, feedback loops, and nonlinearity in general requires accumulation of knowledge over time from thoughtful strategy evaluations under a variety of contexts.

Résumé : Les difficultés de l'évaluation de stratégies d'ensemble complexes découlent des questions de complétude, d'attribution, et de complexité. Dans ce dernier cas, il est nécessaire de porter attention à la synergie entre les interventions, les boucles de rétroaction, et d'autres formes de non-linéarité. Cet article passe en revue et détermine dans quelle mesure six approches différentes à l'évaluation de stratégies réussissent à surmonter ces difficultés, dont une approche mise au point à la lumière de l'étude initiale. Aucune de ces approches ne représente une solution idéale pour ce qui est des questions de complexité. L'approche quantifiée avec modèle logique suggère le besoin de simplifier et

Corresponding author: Robert Schwartz, University of Toronto, Dalla Lana School of Public Health, Ontario Tobacco Research Unit, 155 College St, Ste 530, Toronto, ON M5T 3M7; <Robert.Schwartz@utoronto.ca>

de s'abstenir d'évaluer toutes les chaînes de causalité dans les stratégies complexes. L'analyse de contribution aux voies d'intervention, une approche encore en perfectionnement, explore la possibilité d'utiliser l'analyse de contribution pour valider des propositions d'évaluation découlant de la documentation, de la théorie du programme, et de données d'évaluation incomplètes. En règle générale, une meilleure compréhension de la synergie, des boucles de rétroaction, et des autres formes de non-linéarité exige l'accumulation de connaissances par l'évaluation sérieuse de stratégies dans des contextes variés et sur une longue période.

COMPREHENSIVE AND COMPLEX STRATEGIES

There is broad acceptance of the need for strategies that involve multiple interventions and organizations and whose success depends on synergies, feedback loops, and nonlinear progression. The article presents a critical review of approaches that have been used in evaluating complex tobacco control and other public health strategies with an aim to identifying opportunities for improvement.

Comprehensive strategies pose substantial and interesting challenges to the evaluation endeavour. They involve the weaving together of a variety of policy and program interventions aimed at achieving similar overall objectives. Comprehensive strategies are generally both complicated and complex (Rogers, 2008). They are complicated in that they involve many interventions, players, and locales. Their complexity involves the expectation of synergies, feedback loops, and subsequent nonlinear progression toward the achievement of intermediate and longer-term objectives (Uphoff, 1992).

As governments have become increasingly committed to evaluation and to performance management, they seek to evaluate not only unique interventions but also the effects of comprehensive strategies. Complex strategy evaluations offer opportunities to learn if and how strategy elements are working together to achieve overall targets. They can provide important information to inform ongoing strategy development, including where changes, enhancements, or cuts might be more and less beneficial.

CHALLENGES IN COMPREHENSIVE TOBACCO CONTROL STRATEGIES

There is widespread agreement in the public health policy and research communities that comprehensive strategies involving concur-

rent implementation of a range of interventions provide the greatest opportunity for changing behaviours associated with preventable disease. Health Canada (2002), the Institute of Medicine (National Cancer Policy Board, 2000), the Centers for Disease Control and Prevention (1999), and, most recently, the WHO International Framework Convention on Tobacco Control (Fong et al., 2006) promote comprehensive tobacco control strategies. Intuitively, comprehensive strategies that include a mixture of evidence-based interventions appear likely to yield results greater than the sum of results expected from single interventions. Yet, to date there is little evidence to support these suppositions (Warner, 2006). Moreover, there is little evidence about what level and mix of interventions is required to impact knowledge, attitudes, and behaviours for different subpopulations under a variety of contexts. Despite widespread belief in the importance of comprehensive tobacco control strategies, there is little evaluative research that provides in-depth understanding of the synergistic effects of strategy components on achieving population-level strategy outcomes. This in-depth level of understanding is required to shape strategies that will maximize progress toward eliminating tobacco-caused morbidity and mortality.

Tobacco control strategies involve a mix of a number of policies and programs targeted at a variety of populations for diverse purposes connected with prevention, cessation, and protection. The complexity of both the intervention mixes and the targeted populations makes it difficult to understand the factors that promote and that inhibit changes in outcome measures. Existing macro-surveys and micro-evaluations leave a number of significant gaps in our knowledge, including

1. variance among communities in changes on various outcome measures
2. attribution of changes to particular interventions (i.e., threats to internally valid causal inferences)
3. attribution of changes to different mixes of interventions
4. how contextual factors influence the effectiveness of interventions or mixes of interventions on changes in outcome measures.

The key challenge of tobacco control strategy evaluation is in determining the contributions of interventions, separately and combined, to changes in population-level tobacco control outcomes. Attribution of population-level changes to mixes of interventions is a daunting challenge. Yet this is what is needed to inform decisions about the

future shaping of complex strategies. Such an undertaking requires a tremendous amount of evaluative information about the design, implementation, fidelity, reach, and effects of interventions. A challenge is how to deal with gaps in evaluative information. Synthesis of existing evaluative information from a variety of sources presents additional challenges of its own related to different sources, definitions, and time periods for which information is available. The quality of evaluative information from a range of sources must also be considered.

As part of a process to develop innovative approaches to evaluating a complex tobacco control strategy, we conducted a critical review of approaches used until now in tobacco control and similar public health strategy evaluation work. The critical review assesses strengths and weaknesses of each approach in addressing three dimensions: comprehensiveness, attribution, and complexity. *Comprehensiveness* refers to the extent to which the approach enables assessment of all strategy components and their contribution to overall strategy outcomes. *Attribution* is concerned with explaining how changes in strategy outcomes are a function of interventions. Assessment of the *complexity* dimension focuses on how each approach deals with synergies among interventions, feedback loops, and nonlinearity (Barnes, Matka, & Sullivan, 2003; Sanderson, 2000; Spicer & Smith, 2007). The final section of the article outlines a new approach to complex strategy evaluation that builds on learnings from this review.

In order to retrieve articles/reviews discussing approaches to evaluating public health strategy interventions, a search was conducted in PubMed using terms such as “community level evaluation,” “state level evaluation,” “country level evaluation,” “comparative evaluation,” and “tobacco control review.” The search was limited to evaluations of health promotion programs, with a focus on tobacco control programs. An additional web search for relevant grey literature reviewed the websites of state organizations considered leaders in tobacco control evaluation.

APPROACHES TO EVALUATING COMPLEX COMPREHENSIVE TOBACCO CONTROL STRATEGIES

The literature search identified two broad types of approaches that have been used in evaluating comprehensive strategies: (a) cross-jurisdictional comparisons and (b) comprehensive within jurisdiction approaches. Within each broad type, we review three specific evaluation approaches (Table 1).

Table 1
Approaches to Evaluating Comprehensive Strategies

1. Cross-jurisdictional comparisons
<ul style="list-style-type: none"> a. Community comparisons b. State/regional comparisons c. Country comparisons
2. Comprehensive within jurisdiction approaches
<ul style="list-style-type: none"> a. State strategy evaluations b. Quantified logic models c. Intervention path contribution analysis

CROSS-JURISDICTIONAL COMPARISONS

Community Comparisons

Cross-jurisdictional comparisons seek to use natural or controlled experiments to evaluate the impacts of interventions as compared with control or comparison jurisdictions. Comparative evaluations of health promotion programs at the community level are relatively common (Merzel & D’Affliti, 2003). From the 1970s to the mid 1990s, randomized control trials were relatively common in conducting community health promotion evaluation research. Such trials targeted entire communities and applied social science theory to the development of interventions designed to modify social environment and individual behaviours. Trials generally involved a group of communities separated by definable boundaries and randomly assigned to either receive a standardized intervention or be a control site. Periodic cross-sectional and cohort surveys and other forms of assessment (e.g., monitoring of community legislation, media coverage, health care provider activity) were conducted at regular intervals before, during, and after the intervention in both intervention and control communities. Results from the intervention communities were compared against those in the control site. Examples include:

- The North Karelia Project (1972–1977). A cardiovascular disease prevention program was implemented in the county of North Karelia, Finland, and compared against a single control site (a neighbouring county). The counties were not randomly assigned to intervention or control conditions and the trial itself had low statistical power as there were only two sites (McAlister, Puska, Salonen, Tuomilehto, & Koskela, 1982).

- The Stanford Five-City Project (1980–1986). A comprehensive program of community organization and health education to decrease cardiovascular disease risk factor prevalence, morbidity, and mortality was implemented using six-year education intervention targeting all residents in two intervention communities, compared against three control sites (Fortmann & Varady, 2000).
- The Minnesota Heart Health Program (1980–1993). Six communities in Minnesota were monitored for cardiovascular disease risk factors and disease rates for 10 years prior to intervention. Three communities were randomly assigned to participate in a staggered five-year education intervention designed to promote community change in cardiovascular disease risk factors and related behaviours. The primary outcome measure was net change (that is, change in pooled intervention communities minus change in pooled comparison communities) in awareness, participation, cognitions, behaviours, risk factors, and disease endpoints. Secondary measures included “linkage” between education components and behaviour change and “coincidence” of community change with the staggered program entry (Jacobs et al., 1986).

The Community Intervention Trial for Smoking Cessation (COMMIT) sought to apply the lessons learned from randomized community trials in cardiovascular disease prevention to a comprehensive smoking cessation program (COMMIT Research Group, 1991). Unlike previous randomized control trials, COMMIT had significant statistical power (with 11 matched communities) and targeted a single behaviour—smoking—rather than a group of behaviours (e.g., diet, smoking, exercise).

The randomized control trial design with a standardized protocol and with enough communities to generate sufficient statistical power ostensibly provided opportunity for robust evaluation. A computerized tracking system was used to keep track of intervention activities, their implementation, and their success. The community context was monitored to keep track of any changes in community context that might have affected community attitudes toward tobacco use. And data for the evaluation were collected through a series of surveys administered to a randomly selected segment of the community and to key intermediaries (e.g., physicians, worksites, religious organizations) at various points during the intervention period. Indices were constructed from these data, and scores were then compared between

intervention and comparison communities. Intervention communities were also compared to one another to determine the COMMIT effect in communities of different sizes and socioeconomic and political environments (COMMIT Research Group, 1991).

The COMMIT intervention did not significantly affect the primary outcome measure—quit rates among heavy smokers—where quitting was defined as having stopped smoking and maintained cessation for at least 6 months prior to the end of the trial. For heavy smokers, the average quit rates for intervention and comparison communities were nearly identical (18% versus 18.7%). There was, however, a statistically significant COMMIT effect on quit rates in the light-to-moderate smoking group, with an average of 3% more such smokers quitting in intervention than in comparison communities. The COMMIT Research Group (1995) suggests that the failure of COMMIT to achieve its goal may be due to two notable limitations: (a) the standardized protocol may have constrained communities from undertaking interventions that might have had greater impact (although, as the Research Group acknowledges, community leaders seemed quite satisfied with the protocol); and (b) the protocol did not permit emphasis on some kinds of policy and environmental change that might have been quite effective (e.g., cigarette tax increases).

The COMMIT evaluation did try to address questions of attribution by examining the role of each of four intervention channels in promoting smoking cessation. However, as Lynn, Thompson, and Pechacek (1995) suggest, the evaluators faced the problem of “synergy.” COMMIT interventions were designed to provide synergy between the various activities, so it was difficult, if not impossible, to separate out the contribution to the outcome of various parts of the COMMIT intervention. This approach was further complicated by issues of measurement: only the achievement of process objectives was clearly measurable and it was impossible to assess the interaction between various activities.

After COMMIT, similar community trials occurred, for example, in Minnesota (the Tobacco Policy Options for Prevention program, designed to reduce youth access to tobacco; Forster et al., 1998), Massachusetts (the Worksite Cancer Prevention Intervention, designed to reduce tobacco use and increase the consumption of fruits and vegetables at the workplace; Sorensen et al., 2003), and Stockholm (the Stockholm Diabetes Prevention Program, designed to reduce the prevalence of Type 2 diabetes in Stockholm, Sweden; Andersson, Bjäras, & Östenson, 2002).

Merzel and D’Afflitti (2003) reviewed 32 community randomized control trials and conclude that all, with the exception of those targeted at HIV prevention, have had modest to statistically insignificant effects when compared against control sites. They point to methodological issues (such as low statistical power, design and sampling issues), the influence of secular trends, and limitations in the interventions (such as brevity, emphasis on a standardized protocol preventing more tailored interventions, low levels of community penetration, a heavy focus on changing individual behaviour rather than affecting communities through policy or regulatory change) as possible sources for such disappointing results. However, the fault may be with the evaluations themselves rather than the interventions. Cheadle et al. (2003) note several shortcomings in the use of randomized control trials to evaluate community interventions. Evaluations of community interventions tend to focus on changes in final outcomes rather than intermediate outcomes. Often, the intervention period in community trials is too short to measure changes in final outcomes and attempting to do so may lead to the conclusion that the program had no impact when, in fact, the long-term impact of a sustained intervention may have been positive and significant.

Analysis

Randomized control community comparisons, such as COMMIT, rate fairly high on comprehensiveness as they often include data collection from multiple sources with attention paid to both processes and results. However, they fall short on both attribution and complexity, resulting in challenges in interpreting outcome results. Inadequate knowledge of synergies, feedback loops, and nonlinearity prevent interpreting similar outcomes in experimental and control communities as indicating that the intervention is ineffective. Alternative explanations are that the contexts of a community, the combination of interventions, or the sequencing of interventions were such that synergy did not happen and take-off points were missed. Lack of understanding about tipping points and the relatively short time periods place additional limitations on this approach.

State/Region Comparisons

Cross-state/region comparisons are less common than community comparisons, perhaps because of the inherent methodological challenges of accounting for within-state variation. The evaluation of the American Stop Smoking Intervention Trial (ASSIST) sought to address this challenge through a unique approach.

The evaluation of ASSIST attempted to demonstrate the benefits of state-level comprehensive tobacco control strategies, but with limited success (National Cancer Institute, 2006). ASSIST was an eight-year (1991–1999) demonstration project involving a variety of policy, media, and program interventions in 17 American states (Ockene et al., 1997). ASSIST was initiated to prevent and reduce tobacco use primarily through the application of policy-based approaches to changing the socio-political environment. Primary analysis of the final outcomes showed that ASSIST states exhibited a statistically greater decrease in smoking prevalence for women and lower per capita cigarette consumption rates over the intervention period. However, when these results were adjusted for state conditions, ASSIST status and per capita cigarette consumption were not found to be significantly related (National Cancer Institute, 2006). Extensive between- and within-state variability appears to have overshadowed differences between ASSIST and non-ASSIST states.

The inherent complexity of ASSIST presented a significant challenge to its evaluators, who found that they could not simply document program processes, implementation, and outcomes and then compare ASSIST and non-ASSIST states. Challenges faced by evaluators included contending with the effects of extant state-based tobacco control programs; measuring changes in social, cultural, and political environments; and attributing individual smoking behaviour change to environmental changes caused by ASSIST interventions. They also had to address variations in state conditions and variations in the interventions implemented in each state. To deal with intervention comprehensiveness and complexity, the evaluators developed new methods focused on measuring socio-political change.

The ASSIST evaluation adopted an ecological approach. It used a conceptual model that identified key constructs and developed metrics to measure these constructs. In brief, these metrics are

- *SoTC (Strength of Tobacco Control) Index*. This multi-element indicator allowed the ASSIST evaluators to measure multiple components of state tobacco control efforts (resources for tobacco use prevention and control; capacity and infrastructure to deliver tobacco control efforts; and policy, media, and program services targeted to achieving ASSIST goals).
- *IOI (Initial Outcomes Index)*. This indicator measured policy change, which was the primary focus of ASSIST, and indicated the intensity of states' tobacco control policies.

- *State conditions.* These covariates allowed the evaluators to standardize final results by adjusting for variations in state facilitating conditions and barriers that were beyond the control of ASSIST. This allowed the evaluators to control for state-to-state variations when making comparisons.

The ASSIST Evaluation Team used a broad variety of data collection methods including phone and in-person interviews, a supplement to the census (TUS-CPS), a newspaper clipping service, reviews of tobacco industry documents, literature surveys, and use of existing databases (e.g., State Cancer Legislative Database, or SCLD). Each metric was developed and tested with input from key stakeholders and public health workers. The evaluation was therefore essentially participatory and collaborative in nature. The evaluators also developed hierarchical methods for aggregating survey items into single index measures, thus allowing the various components of each metric to be combined.

Analysis

The ASSIST evaluation included important innovations that are useful for tobacco control strategy evaluation. Like the COMMIT community comparison trial, it scores high on comprehensiveness. ASSIST certainly broke new ground in developing measures of progress (Strength of Tobacco Control) and intermediate outcome measures. These innovations yield greater understanding of the interaction between context and mechanism in the achievement of tobacco control outcomes. Despite the considerable investment in developing new measures and attention to process and intermediate outcome, ASSIST was unable to adequately address how and when synergies, feedback loops, and multiplier effects were realized in intervention states as compared with comparison states. A noted weakness of the ASSIST study is that strength of tobacco control data (SoTC) were collected only once during the intervention period (National Cancer Institute, 2006). Longer periods of time, more measurement, and greater attention to nonlinearity were not present. The focus on between-state comparisons came at the expense of rich understanding of single-state efforts, synergies and feedback loops created or not realized among interventions within a state, and attribution of changes in population-level outcomes to interventions and mixes of interventions.

Country Comparisons

A number of studies have used cross-country comparisons to evaluate tobacco control interventions. These have largely been econometric

studies. A recent and ongoing large project uses a cross-country survey design.

The International Tobacco Control Policy Evaluation Project (ITC) is a unique undertaking designed to evaluate the effects of tobacco control policies through a longitudinal cross-country survey design. The ITC combines four strategies: (a) a quasi-experimental design (in which countries that are exposed to tobacco control programs are compared to those that are not or that have weaker tobacco control programs); (b) longitudinal cohort design (in which individuals are measured on the same key outcome variables at multiple times, before and after the implementation of the program); (c) measurement of policy-specific variables that are theorized to be initially affected after the program is implemented; and (d) measurement of policy-specific variables that have not changed (Thompson et al., 2006).

ITC evaluators developed a conceptual model outlining how tobacco control policies are expected to change outcomes and influence behaviours and through what causal chain this would occur. This model focuses on how policies affect individual behaviour variables, of which there are two classes: policy-specific variables and psychosocial mediators (Fong et al., 2006). Based on this model, ITC evaluators developed a questionnaire that is administered in parallel prospective cohort surveys of representative samples of adult smokers in several countries.

Analysis

The ITC is innovative and tremendously useful for evaluating individual tobacco control policies. It is, however, not a strategy evaluation tool. It does not deal with attribution, synergies, and complexity within a single-country evaluation.

COMPREHENSIVE WITHIN-JURISDICTION APPROACHES

State Strategy Evaluations

A comprehensive review of American state evaluations of tobacco control strategies was conducted by the Ontario Tobacco Research Unit (OTRU; O'Connor, Cohen, & Osterlund, 2001). The review found that states approach evaluation in a similar fashion, addressing implementation and short- and long-term effects. Population impacts are addressed through monitoring, which the authors note does not link program implementation and exposure with program impact.

Analysis

Leaders in tobacco control, including California, Florida, and Massachusetts, have conducted relatively comprehensive strategy evaluations that embed needs assessments, implementation evaluation, and outcome evaluation in a structured evaluation design. Several of these efforts recognize the need to assess synergies among interventions, but have addressed this need in only a rudimentary fashion. For example, evaluations in California have used regression analysis to tease out the additive effects of interventions to prevent initiation of tobacco use. Multiplicative or synergistic effects resulting from interaction were not addressed.

Quantified Logic Models: An Overview

A “quantified logic model” is an evaluation method designed to deal with the difficulty of establishing causality in complex strategies. Developed by the Joint Evaluation Unit during its evaluation of the European Commission’s support to the United Republic of Tanzania, the quantified logic model is less a method by which to establish causality than a process by which stakeholders can develop a profile of their projects in terms of how they *might* contribute to the overarching goals of a strategy. The European Commission Joint Evaluation Unit (2006) calls this a “light study”: rather than determining the *actual* impact of various interventions in a strategy, the study creates a profile of how each intervention is *likely* to contribute to the strategy’s expected impacts. To better explain this method, it may be useful to use the example of the aforementioned evaluation in Tanzania.

In Tanzania, the Joint Evaluation Unit faced what Toulemonde and Carpenter (2011) term “the evaluability barrier”: the impossibility of assessing the achievement of overarching goals in a country-level strategy. The Joint Evaluation Unit sought to determine how all European Commission-funded projects in Tanzania contributed to the overarching goal of poverty reduction. However, the number of cause-and-effect assumptions made by the unit during this evaluation was in excess of one hundred—far beyond the analytic capacity of any evaluation method. Initially, the unit sought to break this evaluability barrier by conducting secondary analysis of existing project evaluations. They were encouraged to do so by the fact that Tanzania is one of the most evaluated countries in the world. However, despite this favourable condition, the unit found that only 20% of the European Commission’s financial support to Tanzania had been covered by

either a project evaluation or a rating of impact. In addition, much available impact information either did not pertain to poverty reduction or consisted solely of guesstimates. The unit was therefore forced to develop an alternate methodology for the evaluation.

The evaluators took their original web of cause-and-effect assumptions and simplified it into a “strategic logic model” with just seven boxes and a dozen arrows. The model was further simplified into five “paths toward poverty reduction.” These paths were (a) conducive development environment, (b) growth of economic activities where the poor engage, (c) reduction of vulnerability, (d) equitable access to basic services, and (e) expanded services while protecting quality (European Commission Joint Evaluation Unit, 2006). The whole portfolio of EC interventions was then analyzed in terms of their potential contribution to the five paths.

Each intervention was assumed to have some potential impact on poverty reduction, but the magnitude and direction of this impact were not considered. Rather, the unit looked at how the potential impact of the intervention was divided among the five paths. Weights were based solely on potential contributions, not on actual contributions. Determining actual contributions would demand in-depth studies. These “light studies” sought only to assess *likely* contributions. Monitoring or evaluation reports for each intervention were used as a basis for these light studies in order to assign weights to each of the five paths. Where such reports did not exist, the unit conducted its own analysis of the intervention to determine the appropriate weights. After all the light studies were completed, they were then weighted by the amount of EC financial support delivered through each intervention. The resulting weighted average was the “quantified logic model.” This model provided an overall picture of EC support in terms of the five paths toward poverty reduction. It allowed the EC to examine how the various interventions it supported (which ranged from promoting equal access to basic education, to building better rural roads, to supporting regional trade negotiations) would likely contribute to the overarching goal of poverty reduction and through which paths this would be achieved.

In conclusion, a quantified logic model is a way through which evaluators can address the evaluability barrier in large, complex evaluations. In such evaluations, interventions are diverse, covering a broad variety of sectors and achieving different outcomes. However, they may be joined by some overarching goal (e.g., poverty reduction). Determining how each intervention contributed to the overarching

goal may just be impossible, but it is still useful to funders to gain a general picture of how each intervention *may* make such a contribution. Funders may also wish to gain a general picture as to how their funding is being allocated across interventions to general paths designed to achieve the overarching goal. For example, in Tanzania, the European Commission may find it useful to find out how much of its financial support is allocated toward the reduction of vulnerability (one of the five paths to poverty reduction). A quantified logic model provides this general picture by creating a profile of each intervention in terms of its potential impact on the overarching goal and by synthesizing these profiles by weighting them based on financial support. Funders can then look at the quantified logic model as a reference for determining how their funds may potentially contribute to the overarching goal(s).

Analysis

The quantified logic model approach contributes an important realistic perspective to the notion of comprehensiveness in evaluating strategies. It provides an overall picture without pretending to have assessed the multitude of causal chains involved in promoting strategy goals. Even in a setting known for its high rate of evaluation coverage, the authors found that only a fraction of the necessary evaluation was in place. The pragmatic approach for overcoming this barrier is innovative and helpful. In a similar fashion, the evaluators provide a picture of how interventions may be affecting long-term overarching outcomes without pretending to precise measurement of attribution. Their point that the general picture is useful for policymakers is noteworthy. The approach does not move us forward in evaluating characteristics of complexity including synergies, feedback loop, and takeoff points.

Merging innovations from the quantified logic model approach with John Mayne's (2001) idea of contribution analysis has resulted in a new model for evaluating comprehensive tobacco control strategies—intervention path contribution analysis.

INTERVENTION PATH CONTRIBUTION ANALYSIS (IPCA)

Building on previous work, we have developed a new approach to complex strategy evaluation that facilitates strategy planning, evaluation planning, and answering central evaluation questions about the overall effectiveness of strategies.

Central to our approach is the identification of evidence-based paths to achieving strategy objectives and evidence-based estimation of the expected and actual contributions of various interventions to progress on each path. Using an innovative intervention path logic model based approach in tandem with contribution analysis, we assess the expected and actual contributions of policy and program interventions to key paths to successful prevention, protection, and cessation as identified in the literature. The comprehensive and integrated strategy evaluation synthesizes evaluative information: literature syntheses, performance measurement, direct evaluation, self-evaluation, and population-level macro monitoring. This section demonstrates how this approach is being applied to evaluating the Smoke-Free Ontario Strategy (SFOS).

The IPCA approach builds on previous work by Douthwaite, Juby, van de Fliert, and Schulz (2006) on path logic models and Toulemonde and Carpenter (2011) on quantified logic models. These models provide useful new ways of describing complex strategies and begin to demonstrate a sense of the roles of various strategy components. However, they are unable to provide a comprehensive and integrated overview of the direct and interactivity effects of program and policy interventions on achieving desired strategy outcomes.

To fill this gap, IPCA relies on contribution analysis (see Mayne, 2001, 2004, 2006) combined with systematic collection and synthesis of evaluative information and evidence from the field and from the literature. IPCA has seven stages:

1. Identify high-level macro outcomes
2. Determine key evidence-based paths to achieving outcomes
3. Identify relevant “interventions”
4. Assess expected contribution of each intervention to each path through literature syntheses
5. Assess actual contributions of each intervention to each path through evaluative information syntheses and contribution analysis
6. Assess contributions of each path to high-level macro outcomes through contribution analysis
7. Assess interactions and synergies through contribution analysis.

Identifying High-Level Macro Outcomes

High-level macro outcome indicators should be obvious from the strategy plan. Where these are not clearly identified by strategy planners, evaluators should seek agreement on appropriate indicators. In order to evaluate a strategy, it is essential that tools are in place to collect data reflecting changes on these indicators over the course of time. Population-level outcome indicators identified for the SFOS include initiation and prevalence of tobacco use among children, youth, and young adults; exposure to second-hand smoke in public spaces, workplaces, and homes; prevalence and number of quit attempts; average cigarette consumption; duration of smoking abstinence among quitters; and tobacco use prevalence. Data for these indicators are collected annually or semi-annually through a battery of provincial and federal population surveys. OTRU analyzes these data and produces annual monitoring reports on progress in achieving these outcomes.

Determining Key Evidence-Based Paths to Achieving Outcomes

Sources for determining key paths include existing documentation and logic models that accompany the strategy and published and grey literature. Key paths should be agreed with key strategy stakeholders. The field of tobacco control has generated large volumes of studies about the influence of interventions. Review and synthesis of this literature reveal the following key paths:

	Prevention	Cessation	Protection
Social Climate	X	X	X
Availability	X	X	X
Knowledge	X	X	X
Awareness		X	X
Social Support	X	X	

Identifying Relevant Interventions

Within each goal area (i.e., prevention, cessation, and protection), all strategy interventions that aim to affect each path should be identified. Existing program logic models and other strategy documents are a prime source of information. It is also useful to discuss with program managers and strategy planners as well as to conduct a critical review of the literature about the intervention. It is important to note interactions among goal areas. For example, the protection

substrategy appears as an intervention for cessation, as restrictions on smoking are known to generate quit attempts.

Assessing Expected Contribution of Each Intervention to Each Path

The expected contribution is determined from systematic literature searches, reviews, and syntheses. In synthesizing evidence, special attention is paid to the particular contexts under which the intervention is being implemented in Ontario. To the extent possible, realist synthesis is being pursued. However, gaps in the literature and time constraints present challenges to the conduct of full-fledged realist syntheses. Expected contributions are categorized as null, small, medium, and large. This component of the complex evaluation includes identifying theories of change both for specific interventions and for recommended combinations of interventions.

Assessing Actual Contributions of Each Intervention to Each Path

All available evaluative information is harnessed to assessing the actual contribution of each intervention to each path. Monthly and annual performance data collected in our web-based performance indicators monitoring system (PIMS) provide information on implementation, including activities, reach, and outputs. Self-evaluations conducted by funded agencies with 10% of their allocations provide information on impacts. OTRU assures the quality of self-evaluations with a formative evaluation support function in which evaluation plans and draft evaluation reports are reviewed through an evaluation review board process. OTRU performs direct evaluations of interventions that span organizational boundaries and that are of particular importance for assessing strategy impacts. These include formative and outcome evaluations of interventions implemented through Ontario's thirty-six Public Health Units. Direct evaluation work also includes cluster evaluations, thematic evaluations, and demonstration projects.

All relevant evaluative information is synthesized to estimate the actual contribution of each intervention to each path at the population level, taking into account both effectiveness and reach. Where insufficient evaluative information is available, further data collection may be initiated, depending on feasibility within time and resource constraints. Actual contribution is expressed as null, small, medium, and large. Where there is insufficient information, this is noted.

The imprecise and tentative nature of this evidence requires further validation. Following Mayne's (2006) model, this validation is sought through key informant interviews, enabling the creation of contribution stories. Mayne recommends surveying knowledgeable others, including experts, program managers, and beneficiaries, about why they think particular interventions are or are not having their desired effects and about whether other programs may be responsible for observed changes.

Assessing Contributions of Each Path to High-Level Macro Outcomes

Once contributions to each path are determined, gaps are revealed in the extent to which paths are being progressed. The picture that emerges demonstrates where there is a need to focus further attention and perhaps where too much attention is being focused.

Assessing Interactions and Synergies

This approach does not allow for formulaic calculations of interactions among interventions or for synergies created. However, attention is given to their existence in comparing actual to expected contributions and in developing contribution stories through key informant interviews.

BENEFITS AND CHALLENGES

IPCA has now been partially applied to the evaluation of the Smoke-Free Ontario Strategy. This has proven to be a rich, yet resource-intensive undertaking. Here we outline the main benefits and challenges encountered in applying IPCA.

The exercise identified substantial gaps both in the strategy itself and in evaluative information to understand the effects of unique interventions and how they interact to create synergies. For example, *Social Climate* was identified as a key path to achieving the desired long-term outcomes of prevention, protection, and cessation. Yet there are no good measures of social climate change at the population level and we know little about if and how interventions in each area are affecting social climate. A key path to achieving long-term prevention outcomes are increasing the *Knowledge* of youth about the harms of tobacco and about the tobacco industry. The evaluation identified five Smoke-Free Ontario interventions that theories of change and

previous literature suggested were likely to affect knowledge: a youth development program, small grants to high schools, an in-class curricular program, a promotional campaign program known as “stupid.ca,” and a comprehensive tobacco control program for university and college campuses. Existing evaluative information confirmed that both the youth development and high school grants programs indeed had positive effects on knowledge among affected youth. However, the reach of these programs is limited. Only two interventions were found to have an expected effect on *Availability* of tobacco products—taxes and restrictions on selling tobacco to minors. While both of these were found to be effective measures, outcome measurement indicates that tobacco is still easily accessible. This finding demonstrates a clear need for additional interventions if the government is serious about preventing youth from smoking initiation.

In many areas, the expected contributions suggest that the family of interventions in place is likely to have an impact on the main paths to achieving primary objectives. However, the literature is generally insufficiently specific about what specific mechanisms need to be in place under a variety of contexts for particular interventions to achieve the desired effect. From the known contribution side, we are often able to assess that the existing interventions are unlikely to have sufficient impact on main paths, owing to their limited reach. Interviews with key informants validate these assessments.

Generating the information to conduct our analyses was more challenging and resource-intensive than expected. It also demonstrated that, generally, in both strategy and intervention design, decisions were not sufficiently rooted in available evidence. Stemming from this is a recommendation that a baseline intervention path analysis be conducted at the strategy design stage, and then used as a basis for building the evaluation framework.

The results of IPCA are extremely useful at the stage of strategy redesign. Despite many gaps in the availability of evaluative information, sufficient data exist to point out where there is a need for greater attention, more resources, new interventions, and expanded reach. With IPCA, substantial progress has been made in strategy evaluation. It provides a very useful evidence-informed assessment of the state of a strategy. Yet to come in this approach is a way to tease out the effects of synergies, feedback loops, and nonlinearity. It may take a number of years of IPCA evaluation implementation on a given strategy before these can be more adequately addressed.

DISCUSSION: IMPLICATIONS FOR THEORY AND PRACTICE

A variety of approaches have been used to evaluate the impacts of comprehensive tobacco control and other public health strategies. Several address the complicated nature of comprehensive strategies by attempting to be comprehensive in addressing multiple interventions, processes, and intermediate and longer-term outcomes. Others utilize the power of experimentation and comparison. However, few approaches address the challenges of assessing the inherent complexity of comprehensive strategies. Neither the cross-jurisdictional comparative approaches nor the within-jurisdiction approaches provide much insight into the interweaving of interventions that results in synergy and feedback loops. Those wishing to develop and refine better strategies therefore have insufficient knowledge as to which interventions to deploy in what sequence and in what combinations under varying contexts.

The quantified logic model approach offers important advances in conceptualizing how to go about making sense of the mess that comprehensive strategies present to evaluators and others. It introduces the notions of light studies, key paths, and simplified quantification. Merging this approach with contribution analysis in intervention path contribution analysis has been an exciting and challenging exercise.

The state of the art in evaluating comprehensive and complex strategies is not yet advanced. Challenges are substantial. This review suggests that evaluators should approach the task with humility, caution, patience, and modesty. Rigorous experimental and quantitative approaches alone cannot meet the challenges of evaluating complex strategies. The material does not allow for regression analyses resulting in simple coefficients that tell us how much an intervention has contributed, how it has acted synergistically, how it has contributed to or been influenced by feedback loops, and how other forms of non-linearity are expressed. Comprehensiveness in collection, analysis, and synthesis of evaluative information is likely a key ingredient for generating useful strategy evaluation. Mapping out expected contributions based on literature and program theory has proven very helpful. Measuring actual contributions is challenging, but should be pursued where possible. Yet evaluators and their clients should recognize that identifying and assessing all causal chains is in the realm of the desired but not achievable. Creating stories based on the best available evaluative information validated through thoughtful contribution analysis offers hope. Over time, the accumulation

of knowledge from multiple strategy evaluations conducted in this way in a variety of contexts will allow for comparisons over time and among jurisdictions, allowing for gradual accumulation of knowledge about synergies, feedback loops, and other forms or nonlinearity in the complex interweaving of interventions that constitute comprehensive strategies.

REFERENCES

- Andersson, C., Björas, G., & Östenson, C. (2002). A stage model for assessing a community-based diabetes prevention program in Sweden. *Health Promotion International, 17*(4), 318–327.
- Barnes, M., Matka, E., & Sullivan, H. (2003). Evidence, understanding and complexity. *Evaluation, 9*(3), 265–284.
- Centres for Disease Control and Prevention. (1999). *Best practices for comprehensive tobacco control programs*. Atlanta, GA: US Department of Health and Human Services, Centres for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
- Cheadle, A., Beery, W., Greenwald, H., Nelson, G., Pearson, D., Senter, S., et al. (2003). Evaluating the California Wellness Foundation's health improvement initiative: A logic model approach. *Health Promotion and Practice, 4*, 146–156.
- COMMIT Research Group. (1991). Community Intervention Trial for Smoking Cessation (COMMIT): Summary of design and intervention. *Journal of the National Cancer Institute, 83*, 1620–1628.
- COMMIT Research Group. (1995). Community Intervention Trial for Smoking Cessation (COMMIT): II. Changes in adult cigarette smoking prevalence. *American Journal of Public Health, 85*, 193–200.
- Douthwaite, B., Juby, T., van de Fliert, V., & Schulz, S. (2006). Impact pathway evaluation: An approach for achieving and attributing impact in complex systems. *Agricultural Systems, 78*(2), 243–265.
- European Commission Joint Evaluation Unit. (2006). *Evaluation of the European Commission's Support to the United Republic of Tanzania*. Brussels, Belgium: Author. Retrieved from http://ec.europa.eu/euro-peon/evaluation/reports/2006/824_vol2.pdf

- Fong, G., Cummings, M., Borland, R., Hastings, G., Hyland, A., Giovino, G., et al. (2006). The conceptual framework of the International Tobacco Control (ITC) Policy Evaluation Project. *Tobacco Control*, 15(suppl. III), iii3–iii311.
- Forster, J. L., Murray, D. M., Wolfson, M., Blaine, T. M., Wagenaar, A. C., & Hennrikus, D. J. (1998). The effects of community policies to reduce youth access to tobacco. *American Journal of Public Health*, 88, 1193–1198.
- Fortmann, S. P., & Varady, A. N. (2000). Effects of community-wide health education program on cardiovascular disease morbidity and mortality: The Stanford Five-City Project. *American Journal of Epidemiology*, 152(4), 316–323.
- Health Canada. (2002). *The Federal Tobacco Control Strategy (FTCS): A framework for action*. Ottawa, ON: Health Canada. Retrieved from http://www.hc-sc.gc.ca/hl-vs/tobac-tabac/about-appropos/role/federal/strateg/index_e.html
- Jacobs, D. R., Luepker, R. V., Mittelmark, M. B., Folsom, A. R., Pirie, P. L., Mascioli, S. R. ... Blackburn, H. (1986). Community-wide prevention strategies: Evaluation design of the Minnesota Heart Health Program. *Journal of Chronic Diseases*, 39(10), 775–788.
- Lynn, W., Thompson, B., & Pechacek, T. (1995). Community intervention trial for smoking cessation: Description and evaluation plan. In *Community-based interventions for smokers: The COMMIT field experience* (Tobacco Control Monograph No. 6, NIH Pub. No. 95-4028, pp. 27–52). Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute.
- Mayne, J. (2001). Addressing attribution through contribution analysis: Using performance measures sensibly. *Canadian Journal of Program Evaluation*, 16(1), 1–24.
- Mayne, J. (2004). Reporting on outcomes: Setting performance expectations and telling performance stories. *Canadian Journal of Program Evaluation*, 19(1), 31–60.
- Mayne, J. (2006). Performance studies: The missing link? *Canadian Journal of Program Evaluation*, 21(2), 201–208.

- McAlister, A., Puska, P., Salonen, J. T., Tuomilehto, J., & Koskela, K. (1982). Theory and action for health promotion: Illustrations from the North Karelia Project. *American Journal of Public Health*, 72, 43–50.
- Merzel, C., & D’Afflitti, J. (2003). Reconsidering community-based health promotion: Promise, performance, and potential. *American Journal of Public Health*, 93(4), 557–574.
- National Cancer Institute. (2006). *Evaluating ASSIST: A blueprint for understanding state-level tobacco control* (Tobacco Control Monograph No. 17). Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. NIH Pub. No. 06-6058. Retrieved from http://cancercontrol.cancer.gov/tcrb/monographs/17/m17_complete.pdf
- National Cancer Policy Board. (2000). *State programs can reduce tobacco use*. Washington, DC: Institute of Medicine.
- Ockene, J. K., Lindsay, E. A., Hymowitz, N., Giffen, C., Purcell, T., Pomrehn, P., et al. (1997). Tobacco control activities of primary-care physicians in the Community Intervention Trial for Smoking Cessation. *Tobacco Control*, 6, 49–56.
- O’Connor, S. C., Cohen, J. E., & Osterlund, K. (2001) *Comprehensive tobacco control programs: A review and synthesis of evaluation strategies in the United States* (Special report). Toronto, ON: Ontario Tobacco Research Unit.
- Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, 14(1), 29–48.
- Sanderson, I. (2000). Evaluation in complex policy systems. *Evaluation*, 6(4), 433–454.
- Sorensen, G., Stoddard, A. M., LaMontagne, A. D., Emmons, K., Hunt, M. K., Youngstrom, R., ... Christiani, D. C. (2003). A comprehensive work-site cancer prevention intervention: Behavior change results from a randomized controlled trial (United States). *Journal of Public Health Policy*, 24(1), 5–25. doi:10.2307/3343174
- Spicer, N., & Smith, P. (2007). Evaluating complex, area-based initiatives in a context of change. *Evaluation*, 14(1), 75–90.

- Thompson, M. E., Fong, G. T., Hammond, D., Boudreau, C., Driezen, P., Hyland, A., ... Laux, F. L. (2006). Methods of the International Tobacco Control (ITC) four country survey. *Tobacco Control, 15*, 12–18.
- Toulemonde, J., Carpenter, D., & Raffier, L. (2011). Coping with the evaluability barrier: Poverty impact of European support at country level. In K. Forss, M. Mara, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 123–144). New Brunswick, NJ: Transaction.
- Uphoff, N. (1992). *Learning from Gal Oya: Possibilities for participatory development and post-Newtonian social science*. Ithaca, NY: Cornell University Press.
- Warner, K. E. (2006). Tobacco policy research: Insights and contributions to public health policy. In K. E. Warner (Ed.), *Tobacco control policy* (pp. 3–86). San Francisco, CA: Jossey-Bass.

Robert Schwartz is Executive Director of the Ontario Tobacco Research Unit and Associate Professor in the Dalla Lana School of Public Health at the University of Toronto. Dr. Schwartz is Editor-in-Chief of the *Canadian Journal of Program Evaluation* and Principal Investigator of the CIHR Strategic Training Program in Public Health Policy. At OTRU, Dr. Schwartz directs a comprehensive evaluation and monitoring program that includes surveillance, monitoring, evaluation, performance measurement, and evaluation support and quality assurance. His research interests include accountability, managerial values and crisis; the politics and quality of evaluation; performance measurement and performance auditing; evaluation of complex strategy initiatives; evaluation of tobacco control programs and policies; and public health policy. He has published widely about accountability, public health policy, policy change, program evaluation, and government–third sector relations.

Gillian Pais has an M.A. in sociology from the University of Chicago and a B.A. (Hons.) in economics and international relations from the University of Toronto. After graduation, she held various positions in the research and public sector, including as a research assistant at the Ontario Tobacco Research Unit, adjunct to the Ontario Ministry of Health Promotion. Gillian is currently an engagement manager at McKinsey & Company, a global management consulting firm.