

## A THEORY-BASED EVALUATION FRAMEWORK FOR PRIMARY CARE: SETTING THE STAGE TO EVALUATE THE “COMPARISON OF MODELS OF PRIMARY HEALTH CARE IN ONTARIO” PROJECT

Margo S. Rowan  
William Hogg  
University of Ottawa  
Élisabeth Bruyère Research Institute  
Ottawa, Ontario

Lise Labrecque  
Élisabeth Bruyère Research Institute  
Ottawa, Ontario

Elizabeth A. Kristjansson  
University of Ottawa  
Ottawa, Ontario

Simone Dahrouge  
Élisabeth Bruyère Research Institute  
Ottawa, Ontario

**Abstract:** Primary care reform has triggered a flood of demonstration projects across Canada that need to be evaluated. This presents a challenge to an evaluator who is uncertain about how to convince clinical investigators to think beyond traditional research designs toward using evaluation approaches. The purpose of this article is to describe the application and benefits of using a theory-based evaluation framework for a large evaluation of four unique models of primary care delivery in Ontario, the Comparison of Models of Primary Health Care in Ontario (COMP-PC) project. Lessons learned are drawn from the authors' experience in applying the theory-based approach, including the benefits and limitations of having a common framework to facilitate model comparison.

**Résumé :** La réforme des soins de santé primaires a déclenché partout au Canada un déferlement de projets expérimentaux que l'on doit maintenant évaluer. Cela représente tout un défi pour une évaluatrice qui se demande comment convaincre les cliniciens chercheurs de dépasser leur méthodologie de recherche tradi-

tionnelle pour adopter des approches évaluatives. Cet article a pour but de décrire les applications et les avantages d'un cadre théorique pour effectuer une évaluation de grande envergure de quatre modèles uniques de prestation de services de soins primaires en Ontario, le projet de comparaison de modèles de soins de santé primaires en Ontario (COMP-PC). Les leçons apprises en expérimentant l'approche théorique comprennent les avantages et les limites d'un cadre commun pour faciliter la comparaison des modèles.

## INTRODUCTION

New primary care models are reforming the way health care is being delivered to patients across Canada. Primary care generally refers to “family doctor-type” services delivered to individuals (Muldoon, Hogg, & Levitt, 2006) and is the “element within primary health care that focuses on health care services, such as health promotion, illness and injury prevention, and the diagnosis and treatment of illness and injury” (Health Canada, 2007). In Ontario, primary care (PC) models have evolved through what Hutchinson, Abelson, and Lavis (2001) refer to as “three waves of development”: the 1970s’ alternative organization and funding models, the 1980s’ expanded range of primary care providers, and the 1990s’ primary care reform demonstration projects. We are now in the fourth wave of development involving the injection of new PC models into the health care system, such as family health networks, family health groups and family health teams. Muldoon, Rowan, Geneau, Hogg, and Coulson (2006) suggest that there are more than 10 different PC models in Ontario, yet these models are poorly understood. It is also unclear why there are so many of them. There is now a critical need to study these new forms of primary care delivery to understand their function, capacity, and niche.

Traditionally, knowledge in clinical areas such as medicine has been sought through research studies where randomized controlled trials (RCTs) are held as the “gold standard” to generate new knowledge for predicting treatment effects. However, while both research and evaluation apply research methods and designs used in health and behavioural sciences (Issel, 2004), evaluation does not routinely aspire to RCT designs for several reasons. Evaluation is interested in understanding what is happening in a specific program rather than attempting to generalize findings more broadly; it focuses on naturally occurring groups and events that are presented rather than controlling the setting and focusing on isolated variables (Levin-Rozalis,

2003). This presents a challenge to an evaluator who needs to convince clinical investigators to think beyond RCTs toward using evaluation approaches that are more appropriate to the context, setting, and objectives when studying alternative PC models. Where to begin? One option is to start with theory.

While much has been written about theory-based evaluations (Chen, 2003; Weiss, 2004), there is a lack of practical advice about its utility, strengths, and limitations (Donaldson, 2003; Weiss, 1997b). The purpose of this article is to describe a theory-based evaluation framework and to demonstrate its utility for focusing and structuring a large-scale evaluation of four PC models in Ontario: the Comparison of Models of Primary Health Care in Ontario (COMP-PC) project. Three interdependent processes are outlined in the application of our theory-based approach, starting with a conceptual framework for primary care, followed by a “generic” theory-based logic model for primary care organizations in Ontario, and a measurement mapping example. We discuss how each level of theory application complements the next and sets the stage for fieldwork and the utility of having a common framework to facilitate the comparison of different PC models. Finally, we extract the lessons learned from this process, as we believe that this represents a substantial contribution to understanding the application of theory-based evaluations.

## THEORY-BASED EVALUATIONS

### Background and Benefits

Much has been written in the evaluation literature regarding the importance of articulating the theory behind a program when planning, implementing, and evaluating programs and services (Chen, 2003; Donaldson, 2003; Issel, 2004; Rossi, Freeman, & Lipsey, 2004; Treasury Board of Canada, 2005; Weiss, 1997b, 2004). Of 22 main evaluation models described by Stufflebeam (2001), program theory-based evaluations stand as the one model grounded in a validated theory about how programs operate within a given context to produce outcomes. While Weiss (2004) argues that evaluators do not agree on the meaning of “theory,” theory-based evaluations have been defined simply as “any evaluation that uses program theory or program logic as its guiding framework” (Davidson, 2004, p. 248).

Articulating a program theory can benefit an evaluation in several ways. For example, Chen and Rossi (1992) suggest that a program theory provides a overall framework or model that describes a pro-

gram and specifies “what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts could be generated” (p. 43). Furthermore, a program theory can help at various stages in the lifecycle of an evaluation: planning and designing an evaluation (Grembowski, 2001; Issel, 2004; Weiss, 2004), conducting the evaluation (Donaldson, 2003; Rogers, Petrosino, Huebner, & Hacs, 2000), seeing effects (Donaldson, 2003; Grembowski, 2001; Issel, 2004; McLaughlin & Jordan, 2004), improving programs (Grembowski, 2001; Issel, 2004), and convincing policy makers of a program’s effects (Weiss, 1995).

### Core Elements of Program Theory

While there is flexibility in describing the core elements of a program theory, most theories focus on how a program runs or operates as well as its mechanisms for achieving an outcome. For example, the program theory described by Chen (2003) comprises an action model describing a set of activities to deliver an intervention to a target group and a change model that is activated when the intervention is delivered. McGilton, Fox, and Sidani’s (2005) program theory model includes inputs (characteristics of clients, interveners, and setting), process (components of the intervention with its specific activities and dosage or the amount, frequency, and duration of the intervention), and outcomes (expected changes resulting from the intervention). Finally, Rossi et al. (2004) describe the general components of program theory as three interrelated areas: the service utilization plan (how and why the program’s intended recipients become engaged with the program and receive sufficient services), the program’s organizational plan (activities to be offered, to whom, and at what volume and frequency, and the human, financial, and physical resources required to offer these activities), and the program’s impact theory (the causal relationship between program activities and produced effects).

### Approaches and Methods for Theory Development

There is no consistent approach to program theory development (Rossi et al., 2004). Chen (1990) describes two main approaches for formulating program theory: stakeholder and social science approaches. The stakeholder-based approach focuses on the ideas, values, and perspectives of different stakeholders in developing a model of program theory and relies on methods such as interviews, focus groups, site visits, and observation. This approach is inherent to Wholey’s (1979) early work on evaluability assessment, designed to clarify a program’s

goals and objectives to determine whether or not a program is evaluable. The rationale for the involvement of stakeholders is that they have intimate knowledge of how the program is intended to work and why, and are often the program funders and will therefore be in the best position to act on the information gleaned from the evaluation.

On the other hand, the social science approach relies on objective, published literature to generate theories about how the program obtains its desired effects (Chen, 1990). This approach is based on the impression that stakeholders' perspectives may "overly emphasize the desirability rather than the plausibility of the program" (p. 67), leading to a theory that is not comprehensive enough to capture the causal processes inherent in the program. Applying this type of model assumes that a substantial body of theoretical and validated knowledge exists about how a program functions along with instruments to assist in studying the program (Stufflebeam, 2001). Methods consist primarily of reviewing program documents and conducting literature reviews of relevant empirical evidence on the program domain.

Chen (1990) also offers a third "blended" approach that includes both stakeholder and social science approaches. While this approach eliminates the reliance on one or the other approach, it is not without limitations such as reconciling multiple and conflicting values. Ultimately the decision to use any one of these approaches lies with a mutual decision among key stakeholders and the evaluators as to what would work best based on their preference and given the project budget and timeline.

### Diagramming the Program Theory

A variety of visual aids can be used to describe and depict the program theory. These include conceptual frameworks, program logic models (PLMs), blueprints, action models, theory-of-change models, report cards, and results chains (Anderson, 2005; Beaulac, Goodine, & Aubry, 2004; Goldman & Schmalz, 2006; Hernandez, 2000; Mayne, 2004; Rush & Ogborne, 1991; Stinchcomb, 2001). In recent years the evaluation literature has been replete with articles about PLMs and their uses in primary care services planning and evaluation (Cioffi, Lichtveld, & Tilson, 2004; Cox, 2000; Dykeman, MacIntosh, Seaman, & Davidson, 2003; Fraser & Hollett, 2003; Goertzen, Hampton, & Jeffery, 2003; Moyer, Verhovsek, & Wilson, 1997; Rowan, 2000; Stinson and Wilkinson, 2004; Watson, Broemeling, Reid, & Black, 2004). A PLM is a diagram of a program's key elements that visually depicts

two main elements of a program. First, it describes what a program is supposed to do (main components, inputs, activities and outputs) and with whom (target groups). Next, a PLM describes what type of initial, intermediate, and long-term outcomes it expects to achieve. Usually shown in a flowchart format, a PLM becomes a visual picture of the “logic” or theory behind a program.

PLMs can range from micro (individual program) level to meso (organizational) and macro (system) level. For example, Watson et al. (2004) developed a macro-level, results-based PLM for the entire primary care system, and Watson (2007) developed another results-based PLM for Ontario community health centres (CHCs). These macro-level PLMs summarize the key components needed to measure both the efficiency (process) and the effectiveness (outcomes) of the PC system and CHC program in Ontario. They also delineate the components within the control of the PC system and those that are influenced to a lesser and greater degree by factors external to the organization.

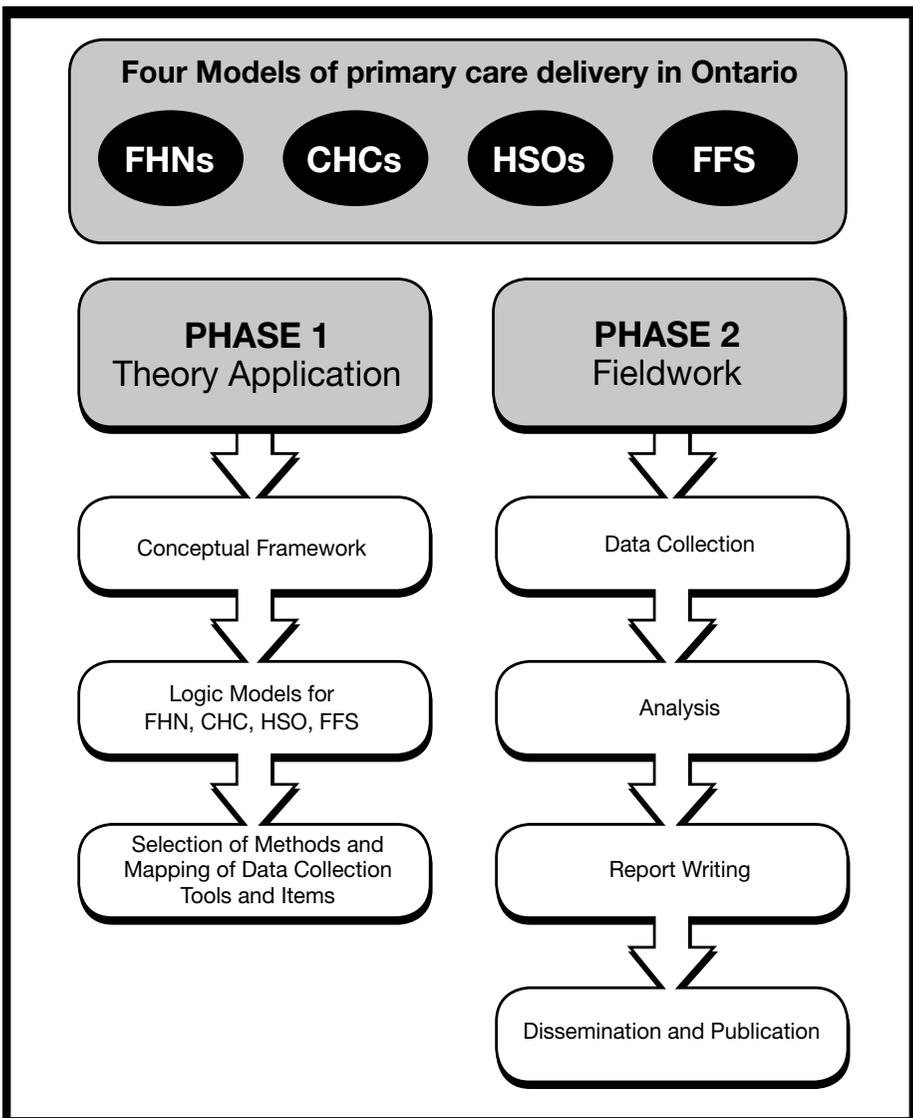
By clearly articulating the implementation steps of a program as well as the expected performance of a program (outcomes), we see the causal linkages between implementation activities and anticipated impact on those participating in and receiving the said activities, thus making PLMs an extremely valuable aid in illuminating a program’s theory of change.

## OVERVIEW OF THEORY DEVELOPMENT AND APPLICATION FOR THE COMP-PC PROJECT

We used a blended approach of stakeholder and social sciences approaches to develop the theory behind four types of PC models in Ontario of interest to the COMP-PC project: Community Health Centers (CHCs), Family Health Networks (FHNs), Health Service Organizations (HSOs), and Fee for Service (FFS). The evaluation team was purposefully selected to provide expertise in evaluation, measurement, and primary health care. It comprised two health program evaluators, an academic family physician, a social scientist, and an epidemiologist. The program evaluators led and educated the team in terms of the theory-based evaluation approach. The academic family physician was a principal investigator, and the social scientist was a co-investigator with the COMP-PC project. They provided valuable insights into the inner workings of the PC models. The epidemiologist provided direction on the project’s data collection tools and completed the measurement map.

Together we used program documentation and narrative literature review (synthesis of existing literature rather than a systematic review), expert consultation, and regular group meetings to develop the theory. As seen in Figure 1, the main objective of our work was to develop a theory base for the four PC models under review (Phase

**Figure 1**  
**Theory-Based Evaluation Process for the COMP-PC Project**



1). This theory would serve as a guide to the field work (Phase 2). Our theory development and application involved the production of (a) a conceptual framework, (b) PLM(s), and (c) a measurement map. Each of these interrelated products and processes is discussed in the following sections. This article focuses on theory development and does not examine the methods, sampling, or analysis from Phase 2 as this is the focus of other upcoming publications (Darouge et al., in press).

### Conceptual Framework

The first step in a theory-driven program evaluation is to develop a conceptual framework (Donaldson, 2003) that “identifies a set of variables and the relations among them that are presumed to account for a set of phenomena” (Carpiano & Daley, 2006, p. 565). We developed our conceptual framework at the onset of our study to help identify key constructs for evaluation, particularly given the complexity and multidimensional nature of primary care (Starfield, 1998). Its development and use were also a means of reducing bias by providing a comprehensive and methodical basis for choosing indicators that cut across all PC models rather than selecting only certain indicators known to highlight the best features of any given PC model.

We began by designing a preliminary draft of our conceptual framework based on previous North American work (Canadian Institute for Health Information, 2003; Committee on the Future of Primary Care, Institute of Medicine, 1994; Division of Health Manpower & Resources Development, Institute of Medicine, 1978; Donaldson, Yordy, Lohr, & Vanselow, 1996; Starfield, 1998). The draft framework was expanded and refined through several concurrent processes:

- Consultation with a group of experts in primary care about the validity of the conceptual framework. At the same time this group was developing definitions for dimensions of primary care through a modified Delphi process and a face-to-face meeting (Haggerty et al., 2007). The group helped to provide direction on the components and subcomponents of the framework and the terms used.
- Search of MEDLINE and the Cochrane Database for English-language literature from January 1995 to January 2006 using the key words “conceptual framework,” “analytical framework,” “primary health care,” “primary care,” “delivery of health care,” “quality of health care,” and “organizational theories.”

- Comprehensive literature review on each of the major dimensions from the original framework (e.g., access, comprehensiveness) and review of documents published by Health Canada and our funder (the Ontario Ministry of Health and Long-Term Care), as well as current texts on primary care and organizational management.
- Evaluation team meetings, including a review of the most recent iteration of the conceptual framework and definitions, presentations by group members on aspects of the framework, and discussions leading to consensus on elements and structure of the framework and linked definitions. Following each meeting, a revised version of the framework was constructed, and the process began again.

The conceptual framework is outlined in Figure 2 (Hogg, Rowan, Russell, Geneau, & Muldoon, 2008). We divided the framework into two main areas, one focused on structure and the other on performance. Structural features allowed us to consider the types of variation in the functioning and organization of the PC models. The structural domain includes three main components: (a) health care system, which describes the policies, stakeholders (e.g., public agencies and professional associations), and factors at the system level that can influence primary care organizations and providers, such as governance and accountability; (b) practice context, which includes the factors at the community level that can influence the organization of the practice and the delivery of care, such as population characteristics; and (c) organization of the practice, which is influenced by its internal structures and processes, such as health human resources.

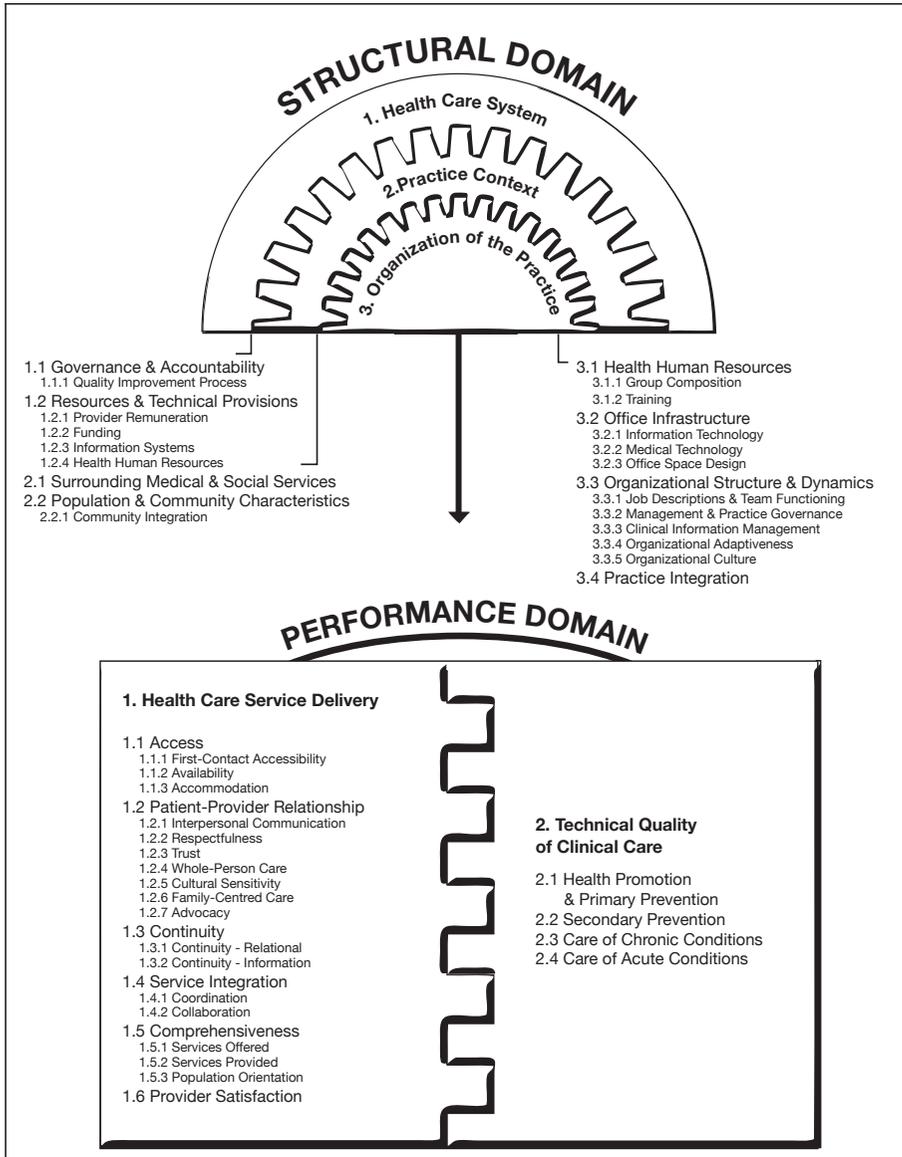
The performance domain focuses on differences in health care service delivery and clinical care provided. It includes cardinal attributes and functions of health care service delivery, such as access, continuity, and the patient–provider relationship, along with the basic features of clinical care, such as management of chronic conditions and health promotion. When combined with performance outcomes, process features helped to identify which factors might contribute to a PC model’s success or deficiency in a certain area.

### Program Logic Model Development

The next step was to apply our conceptual framework to the development of four PLMs representing each type of PC model of interest to the COMP-PC project. These PLMs provided a detailed “picture” of

what the PC models might actually look like in practice and identified important areas to measure, thus moving the conceptual framework closer to the measurement mapping. In writing on her theory-of-change model, Anderson (2005) outlines the differences between a

**Figure 2**  
**Conceptual Framework for Primary Care Organizations**

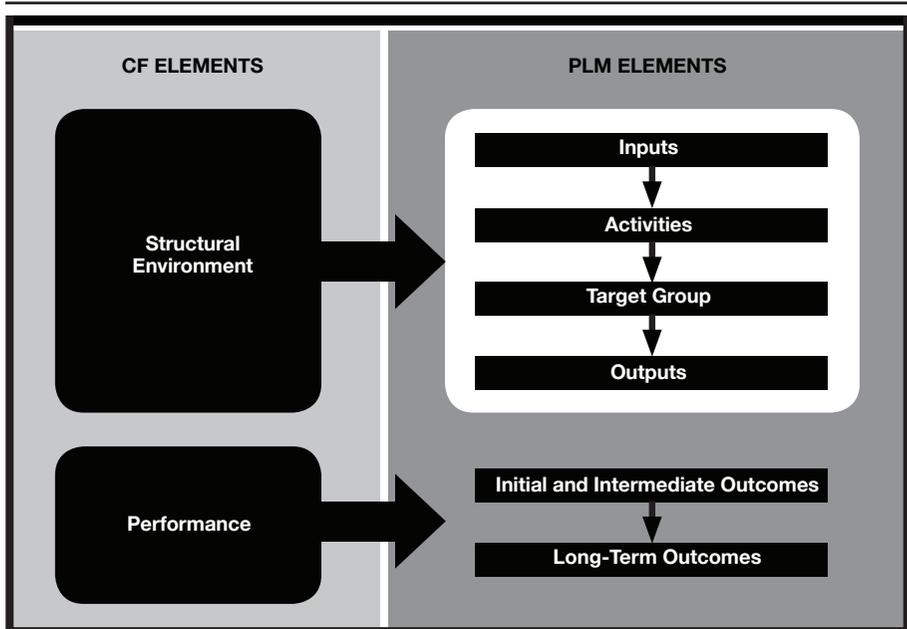


program's theoretical assumptions and its PLM. She states that the theory-of-change model “summarizes work at a strategic level, while a logic model would be used to illustrate the tactical, or program-level, understanding of the change process” (p. 19). So while elements of our conceptual model would apply to all four PC models, the PLMs bring to the surface important differences in the application of these core elements, thus permitting cross-model comparisons.

The link between a conceptual framework and a PLM is illustrated in Figure 3. While the conceptual framework describes the structure and performance domains of primary care more broadly, the PLMs differentiate among PC models. The structural domain of the conceptual framework relates to the inputs, activities, target group, and outputs of the PLM. The performance domain of the conceptual framework relates to the initial, intermediate, and long-term outcomes of the PLM.

Detailed macro-level PLMs for each of the four PC models were developed iteratively between July 2004 and September 2005. There were several steps in their development. First, the conceptual framework

**Figure 3**  
**Link Between the Conceptual Framework (CF) and Theory-Based Program Logic Model (PLM) Elements**



acted as the key conceptual template, providing the focus and content for the PLMs. The conceptual framework was near completion when the PLM work began, so the process was partially simultaneous as well as iterative. Next, the finer details within each PLM were completed based on a review of the literature on various PC models in Ontario commissioned by the COMP-PC investigators (Coulson, 2005; Muldoon, Rowan, et al., 2006). Finally, several drafts of the PLMs were developed and modified by the current authors based on minor changes to the conceptual framework. In the interest of being parsimonious, we present one generic, macro-level PLM for PC in Ontario (see Figure 4).

The use of PLMs in this project helped bring focus to cross-comparisons between and among PC models. The PC models are comparable in many of their performance features, such as providing acute, chronic, and preventative care. They also have similar intended outcomes such as delivering care that is accessible, comprehensive, and integrated. However, the PC models differ in a number of ways. For example, in FHN and CHC models, there are significant differences within the structural environment that would lead us to predict that particular performance features might be different between these models. For example, CHCs have a team of different health professionals, including physicians, nurses, and allied health professionals, to address a wide range of patient issues in one primary care organization. These factors could lead to better performance of CHCs in service integration and comprehensiveness of care. On the other hand, FHNs might perform better than some CHCs in certain aspects of access and continuity because they provide after-hours and telephone health advisory services, enabling clients or patients to easily initiate contact with a provider for a new or existing health problem.

Our use of the theory-based approach to highlight the similarities and differences among PC models is consistent with the work of McGilton et al. (2005). In comparing and contrasting different programs, these authors state that “the application of the theory-driven approach to evaluation requires evaluators to ... document variability in implementation of the intervention, and ... account for variability in the characteristics of clients, interveners, and settings” (p. 32), as opposed to a randomized controlled trial that attempts to control contextual factors that may influence program delivery and outcomes. A theory-driven evaluation “calls for the evaluator to develop a framework that identifies and delineates the relationships amongst the factors that are believed to influence the delivery of the intervention and the achievement of its outcomes” (p. 32). More specifically, McGilton et

al. (2005) look at three factors within programs—the patients, the intervener (staff), and the setting—and they contrast and compare the similarities and differences within each factor and the impact that it can have on a program or intervention. The development of a conceptual framework in the COMP-PC project also allowed for a similar process, one of comparing similarities and differences between PC models and their contextual realities.

### Measurement Mapping

A measurement map takes the theory application one step further to the point of outlining how the main components will be measured. Its purpose is to help guide the evaluation from theory to an examination of the practicality of measuring the effectiveness of various PC models. These types of grids have been used in previous studies of primary care reform and evaluation (Rowan, 2000) and are found within the performance measurement specifications of the federal government's Results-based Management and Accountability Frameworks (RMAFs) (Treasury Board of Canada, 2005). The purpose of our measurement maps was to fit the main elements of the conceptual framework with key areas to measure while considering main differences among the PC models and adding items within the measurement grid to examine those differences.

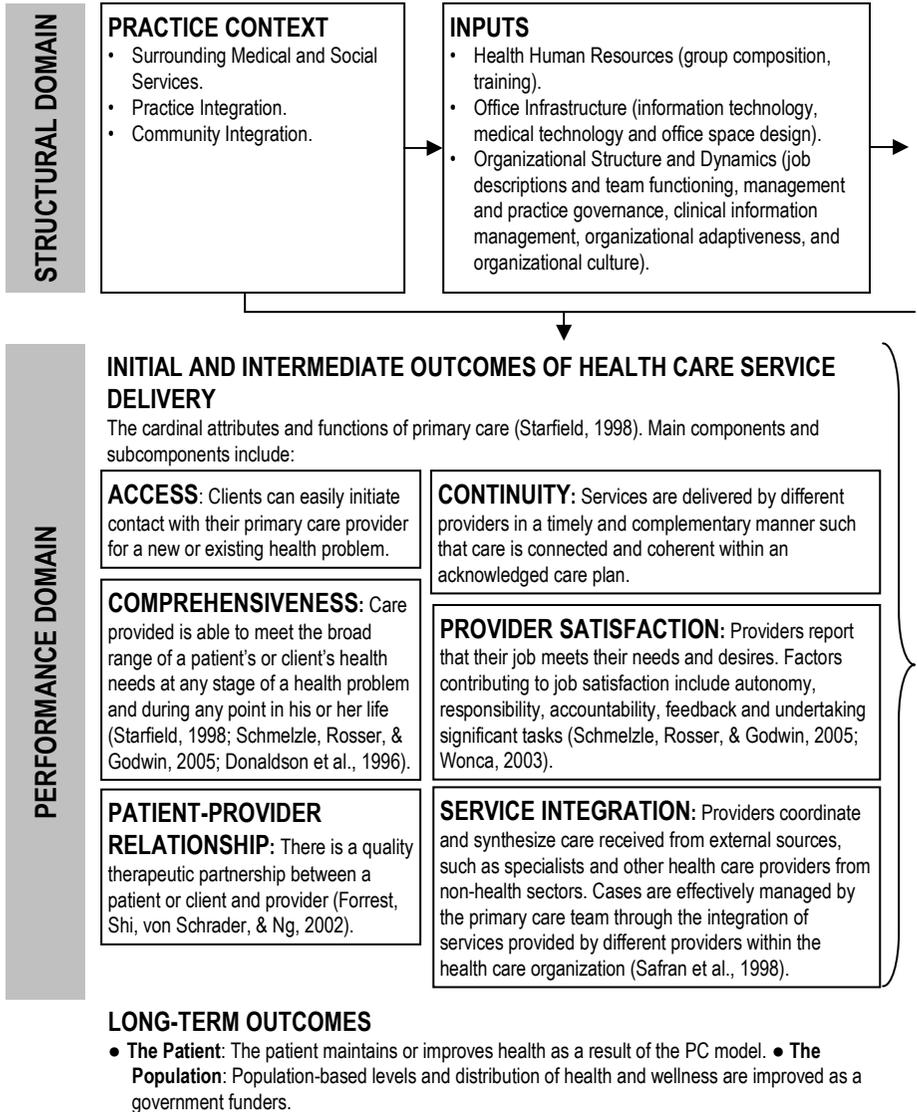
The measurement map was initially developed by an evaluator on the team. She took sections from the conceptual framework and matched each component and subcomponent to the “best fit item” under consideration within the project's data collection tools. The main quantitative measurement tool was the Primary Care Assessment Tool (PCAT) (Shi, Starfield, & Xu, 2001), selected for the project because of its high congruency to concepts in our conceptual framework. Furthermore, the PCAT is measured from the perspective of both patients and providers. When our mapping revealed deficiencies in the content of this instrument, we supplemented it with questions from other sources. For example, we used questions from the National Physician Survey (College of Family Physicians of Canada, 2004) to capture physician satisfaction with various dimensions of their work.

We used qualitative interviews to inform subcomponents of the structural domain of the conceptual framework and to provide information about unanticipated outcomes or unintended effects of the PC models. Interview items were also used to obtain information about the differences in the structural domains among the PC models as gleaned

Figure 4  
Theory-Based Logic Model for Primary Care Organizations in Ontario

## Theory-based Logic Model for Primary Care Organizations in Ontario

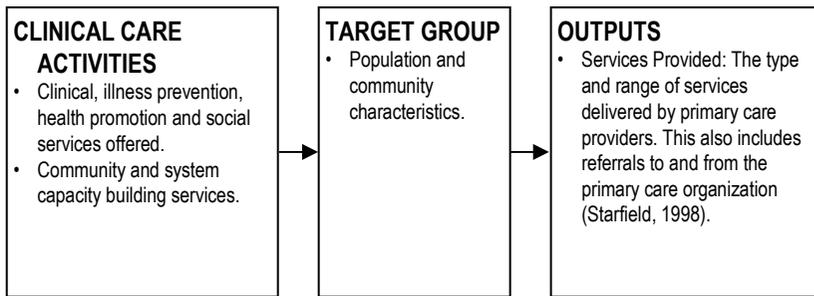
**HEALTH CARE SYSTEM** - The policies, stakeholders (e.g., public agencies, professional associations) and Accountability and Resources and Technical Provisions are two main components here.



**Figure 4**  
**Theory-Based Logic Model for Primary Care Organizations in Ontario (continued)**

**Theory-based Logic Model for  
 Primary Care Organizations in Ontario**

factors at the system-level that can influence primary care (PC) organizations and providers. Governance and



**INITIAL AND INTERMEDIATE OUTCOMES OF TECHNICAL QUALITY OF CLINICAL CARE**

Clinical procedures reflect current research evidence and/or meet commonly accepted standards for technical content or skill (Haggerty et al., 2007). Main components include:

**HEALTH PROMOTION and PRIMARY PREVENTION:** Health promotion is the process of enabling people to increase control over, and to improve, their health (WHO, 1986). Primary prevention is directed towards preventing the initial occurrence of a disorder (WHO, 2005).

**SECONDARY PREVENTION:** Early detection of disease and treatment that may accompany screening (Shah, 2003).

**CARE OF CHRONIC CONDITIONS:** Care for a condition or conditions that are of a chronic nature (Donaldson et al., 1996).

**CARE OF ACUTE CONDITIONS:** Care for health concerns of recent onset, which may range from an acute, relatively minor, self-limited illness, to a complex set of symptoms that could be life threatening, to a mental problem (Donaldson et al., 1996).

**Providers:** The PC workforce functions optimally in providing high quality care to patients. ● **The result of the PC model.** ● **The Payer:** The PC model is economically efficient and sustainable to

from the four separate PLMs. These questions pertained to funding or how a practice was organized and what effect that might have on outcomes related to health care service delivery or technical quality of clinical care. For example, they were asked, “How would you compare a typical day of work in this PC model to a typical day of work in an FFS practice? What is different? Why?” and “How does the presence of other professionals in this PC model affect your work? From your perspective, how is collaborative care influencing the health care experience of patients?” Answers to these questions would lead to exploring care processes and how they might affect certain outcomes such as access, continuity of care, provider satisfaction, and other outcomes seen in the PLM.

After the initial mapping, the evaluation team members met to discuss the appropriate fit of the items with components and subcomponents of the conceptual framework until consensus was reached after three iterations. All performance measurement scales were readily mapped to the appropriate components. An example of a measurement map is given in Table 1.

**Table 1**  
**Measurement Map for COMP-PC Project (excerpt)**

Conceptual framework component	Sample methods, tools, and items				
	Health care service delivery	Patient survey	Provider survey	Practice survey	Chart audit
Access: First contact accessibility	<p><b>Q15.</b> When your provider's office is open and you get sick, would someone from this office see you the same day?</p> <p><b>P116.</b> When your provider's office is open, can you get advice quickly over the phone? If you need it?</p> <p><b>P117.</b> When your provider's office is closed, is there a phone number you can call if you get sick?</p>	<p><b>Q1.</b> When your practice is open and a patient gets sick, would someone from your practice/network see them that day?</p> <p><b>Q2.</b> When your practice is open, can patients get advice quickly over the phone when they think they need it?</p> <p><b>Q3.</b> When your practice is closed, do you have a phone number (other than the Ontario Telephone Health Advisory System) patients can call when they get sick?</p>			

Service  
integration:  
Coordination

**Q14.** When you want to see a specialist, do you get a referral from your provider?

**Q8.** Does your practice phone or send patients the results of lab tests?

**Q9.** Do you think you know about all the visits that your patients make to specialists or special services?

**Q10.** When a patient needs a referral, do you discuss different places the patient might go to get help with their problem?

**A.** Is there a problem (or diagnosis) list in the patient's records? Was problem listed?

**B.** Is there a medication list in the patient's records?

Compre-  
hensiveness:  
Population  
Orientation

**Q26.** Does anyone at your provider's office ever make home visits?

**Q27.** Does your provider know about the important health problems of your neighbourhood?

**Q28.** Does your provider get opinions and ideas from people that will help to provide better health care?

**Q23.** The following questions are regarding community orientation:

- a. Do you make home visits?
- b. Do you think you have adequate knowledge about the health problems of the community you serve?
- c. Do you get opinions and ideas from people that might help to provide better health care?

**Q6.** Does your practice site use the following types of **data** to determine what programs/ services are needed by the communities you serve?

**Q7.** Does your practice site use the following **methods** to monitor and/ or evaluate the effectiveness of services/programs it offers?

**Q8.** Does your practice site use any of the following activities to reach out to the population in the community you serve?

---

As a result of this exercise we had a measurement protocol in place that accounted for most of the structural and performance domains of the conceptual framework. While most of the areas that we were interested in were covered in the mapping exercise, a few areas, such as whole-person care, were left unfilled. These were not a measurement priority, but they have been noted and could be the focus of future item development. Finally, the measurement map also served to guide the evaluators in their analyses. Scales measuring performance outcomes were readily found in the appropriate grid, and structural features that might be of interest in multivariate analyses were readily found under the section of interest. With the measurement

mapping completed, our theory development phase was finished, and we launched into the fieldwork of the project, or Phase 2.

## DISCUSSION

We now turn to a discussion of the “lessons learned for evaluators” in using a theory-based approach to evaluation, based on the evaluation team’s experience.

### 1. Theory-Based Evaluations Are Important to Use, But Not Always

Donaldson (2003) suggests inadequate program conceptualization is one of five key problems that must be overcome to improve “social programming.” Other researchers point to the patent value of using theory-based evaluation approaches (Chen, 2003; Chen & Rossi, 1989; Stinchcomb, 2001) that move beyond a traditional method-driven approach (Donaldson, 2003) that operates more in an unanchored “conceptual vacuum” (Stinchcomb, 2001). However, attention must be given to the conditions under which theory-based evaluations are most appropriately used. In our experience, articulating a program’s theory of change takes considerable time and resources, making this approach prohibitive to evaluations on a tight timeline and budget and particularly difficult to replicate owing to the cost and time involved in theory development.

Leeuw (2003) cites three circumstances under which theory-based evaluation is worth the effort. The first is when a program is expected to have a large impact, such as a province-wide universal influenza vaccine program or diabetic screening and counselling programs. The second is when there are large risks or uncertainty about a program’s impact and unintended side effects, for example, new and innovative programs to prevent substance abuse among youth. Finally, evaluation is vital when a large sum of funding has been committed to a program requiring firm accountability mechanisms, such as province-wide telehealth programs.

Our study about the different PC models met most if not all of the conditions for applying a theory-based evaluation approach. We had a reasonably long time frame (two years) to plan and conduct the study. Our budget for this study was also quite substantial (\$2.4 million), with up to 25 staff members, necessitating tight accountability mechanisms. Phase 1 cost approximately \$100,000 and required over a year of frequent interaction and a significant team effort to

create. There were several reasons why it took so long to complete this phase. As already stated, the project was large, and primary care is a complex area of medicine with dozens of components and subcomponents with which to familiarize ourselves when developing the conceptual framework. Different project members joined partway through the development phase, providing new perspectives and insights that required us to revisit our framework. Also, the concept of evaluation was relatively new to the physicians leading the project, as they came from a more RCT-dominated research design world. Over time they were guided and to a certain degree convinced about the usefulness and appropriateness of evaluation designs. Certainly the use of theory (as in theory-based evaluations) helped us to assure these researchers that evaluation can have an empirical base, thus providing a valuable means to connect their world of empirically based RCT medical evidence with real-world health service delivery evaluations intended to make more informed policies and decisions about models of primary care.

Another condition that made our study suitable for a theory-based evaluation was the assumption that reforms to primary care would have a large impact on the health of Canadians, as primary care is the “foundation of the Canadian health care system” (Canadian Medical Association, 1994, p. 14). According to Romanow (2002), the most important work in improving health care for Canadians is in every interaction between a patient and his or her health care provider. Finally, there were significant risks or uncertainty regarding the impact and unintended effects of PC models, particularly newer models, because these reforms have been made in the absence of robust evidence of what works and why (Muldoon, Rowan, et al., 2006).

## 2. Moving from Conceptual Framework to Program Logic Model to Measurement Mapping Is Not a Direct Fit

While our theory-based approach acted as a sound “foundation” or starting point in a large evaluation effort involving close to 140 primary care sites across Ontario, we encountered some difficulty when trying to “fit” the concepts of the conceptual framework into the structure of a PLM. For example, as demonstrated in Figure 3, at first glance one may see the link between the conceptual framework and the PLM, particularly at a broad-based level. However, in reality, trying to fit the details of the conceptual framework into the PLM did not always produce a direct, one-to-one correspondence

between the two. Thus, judgement was used in deciding where to fit some elements. This is illustrated in the conceptual framework sub-component “comprehensiveness,” which has three elements: services offered, services provided, and population orientation. Although these elements fit well together under the category of comprehensiveness in the conceptual framework, in the PLM the three elements were moved to three different components: services offered were specified under clinical activities, services provided under outputs, and population orientation as a subcomponent of comprehensiveness under outcomes.

### 3. The Theory-Driven Approach Does Not Solve All the Problems Related to Measurement in an Evaluation

Our approach helped to reduce bias in the identification of what gets measured by providing a comprehensive and balanced set of dimensions from which to select methods, tools, and items for measuring the structure and performance of the PC models under study. This is in line with Donaldson’s (2003) idea that theory-driven approaches allow the program evaluator to be “method neutral” by selecting methods guided by the “super ordinate” vision of the program rather than preselecting quantitative, qualitative, or mixed methods. This allowed us to produce measurements to answer the questions “What is working?”, “Why or why not?”, and “How does one PC model compare to another on a particular dimension?” Finally, the theory-based approach helped not only to set the stage for data collection but also to organize our approach to univariate and multivariate analysis, reporting to decision-makers and policy-makers, and publishing in peer-reviewed journals. As Weiss (1997b) suggests, theory-based evaluations can help evaluators make informed choices about what to study and consequently what to analyse and report: “With data that follow the chains of program theory, they can say more useful, relevant, and practical things” (p. 518).

While the theory-driven approach helped to structure and facilitate the selection of methods and mapping of data collection tools and items, it did not solve all the problems associated with our measurement phase. We had to make several compromises owing to limited timelines, concern over participant burden, and lack of existing measurement tools for certain areas. For example, equity or availability were not measures captured by our quantitative tool. However, the theory-based approach allowed us to identify these gaps and adapt our qualitative interview guides to address these.

#### 4. The Extent and Type of Stakeholder Involvement and Weight of Their Input in Theory-Based Evaluation Frameworks Is Not Clear

As we have suggested, there is no one approach to theory-based evaluations. Although we used a blended method of both stakeholder and social science approaches, we relied most heavily on the latter because of the abundance of literature in the primary care field, the lack of preference on the part of our funder as to the type of approach to use, and the fact that a stakeholder-based approach already existed for primary care (Watson et al., 2004). Given more time we would have expanded our stakeholder base to bring together representations from various primary care organizations across Ontario and government health care policy- and decision-makers for an informed discussion about the applicability and utility of the PLMs.

Clearly the research suggests that the validity of the final product of an evaluation is improved by the inclusion of relevant stakeholders in the process, as seen in applications of participatory evaluation approaches (Cousins & Whitmore, 1998). However, despite the benefits of involving stakeholders, the literature offers little guidance about the extent and type of stakeholder involvement in theory development. Nor is it clear how to weight stakeholders' input compared to that of empirical evidence. Issues about stakeholder involvement in theory-based evaluation need to be empirically tested before making sound recommendations. We need to investigate: What does it mean to have stakeholder involvement in theory-based evaluation frameworks? Who should we involve and at what point in the theory development process: research experts? practitioners? patients or clients? political stakeholders? What emphasis should be placed on stakeholders' views compared to evidence-based research findings?

#### 5. The Theory Behind Theory-Driven Approaches Needs Revisiting and Testing

Finally, our conceptual framework will naturally change over time as primary care evolves due to political decisions, learned experiences from providers, demands of patients or consumers, new empirical evidence about primary care delivery, and other contextual factors. We suggest revisiting any theory-based model periodically (e.g., yearly) to maintain its relevance. Furthermore, it would be beneficial to extend our theory-based evaluation process by testing the "model in use" compared to our theoretical model. The PLM could be compared to our data collected in the field to see whether the proposed models

are a true reflection of the actual models. If not, necessary changes could be made.

## CONCLUSION

Evaluation is a critical element of primary care reform. Theory-based approaches anchor evaluation efforts by joining theoretical concepts and operational practices. The use of such approaches can enhance both the theoretical conceptualization of primary care and the practicalities of planning evaluations of primary care reform. Our article reveals the benefits of using this approach when planning an evaluation of a large, complex and multi-faceted delivery model, such as PC. Indeed the theory-based evaluation approach provided a solid platform from which we as evaluators could extend our thinking, planning, and communications. Most importantly, it facilitated the development of an integrated, common PLM and in doing so helped to bring focus to the similarities and differences between and among PC models. From a practical point of view, managers working under various PC models could use our PLM as a springboard for developing evaluation plans, and funders could use it as a discussion piece to develop relevant measurement indicators for community-based programs. Finally, we believe it is possible to transfer this approach to other sectors, such as education, social services, and urban planning, and encourage others to take a closer look at applying and refining this approach.

## ACKNOWLEDGEMENTS

We would like to acknowledge Drs. Laura Muldoon, Robert Geneau, and Grant Russell for their substantial contributions to the final version of the conceptual framework underpinning the Models Project. We thank Gloria Baker for her helpful editing and Guy Mercier for his graphical work. Funding for this research was provided by the Ontario Ministry of Health & Long-Term Care Primary Health Care Transition Fund. The views expressed in this report are the views of the authors and do not necessarily reflect those of the Ontario Ministry of Health & Long-Term Care.

## REFERENCES

- Anderson, A. (2005). An introduction to theory of change. *Evaluation Exchange*, 11(2), 12, 19. Retrieved February 17, 2006, from <<http://www.gse.harvard.edu/hfrp/eval/>>

- Beaulac, J., Goodine, L., & Aubry, T. (2004). *Report card methodology and indicators: Development of the report card on homelessness in Canada. Part 1 – Review of the literature*. Ottawa: Centre for Research in Community Services, Faculty of Social Sciences, University of Ottawa.
- Canadian Institute for Health Information. (2003). *Health care in Canada 2003*. Toronto: Author. Retrieved July 20, 2006, from <[http://secure.cihi.ca/cihiweb/disPage.jsp?cw\\_page=download\\_form\\_e&cw\\_sku=03HCR\(HC\)PDF&cw\\_ctt=1&cw\\_dform=N](http://secure.cihi.ca/cihiweb/disPage.jsp?cw_page=download_form_e&cw_sku=03HCR(HC)PDF&cw_ctt=1&cw_dform=N)>
- Canadian Medical Association. (1994). *Strengthening the foundation: The role of the physician in primary health care in Canada*. Ottawa: Author.
- Carpiano, R.M., & Daley, D.M. (2006). A guide and glossary on postpositivist theory building for population health. *Journal of Epidemiology and Community Health*, 60, 564–570.
- Chen, H.T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Chen, H.T. (2003). Theory-driven approach for facilitation of planning health promotion or other programs. *Canadian Journal of Program Evaluation*, 18(2), 91–113.
- Chen, H.T., & Rossi, P.H. (1989). Issues in the theory-driven perspective. *Evaluation and Program Planning*, 12, 299–306.
- Chen, H.T., & Rossi, P.H. (Eds.). (1992). *Using theory to improve program and policy evaluations*. Westport Park, CA: Sage.
- Cioffi, J.P., Lichtveld, M.Y., & Tilson, H. (2004). A research agenda for public health workforce development. *Journal of Public Health Management and Practice*, 10(3), 186–192.
- College of Family Physicians of Canada. (2004). *National physician survey*. Retrieved October 10, 2005, from <[http:// www.nationalphysician survey.ca](http://www.nationalphysician survey.ca)>
- Committee on the Future of Primary Care, Institute of Medicine. (1994). *Defining primary care: An interim report*. Washington: National Academy Press.
- Coulson, D.B. (2005). *Primary health care models in Ontario*. Toronto: Coulson & Associates.
- Cousins, J.B., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory*

- evaluation. New Directions for Evaluation, 80* (pp. 5–24). San Francisco: Jossey-Bass.
- Cox, R.J. (2000). Using programme logic models in evaluation: A review of the literature and the spinal outreach team experience. *British Journal of Occupational Therapy, 63*, 115–120.
- Dahrouge, S., Hogg, W., Russell, G., Geneau, R., Kristjansson, E., Muldoon, L., et al. (in press). The Comparison of Models of Primary Care in Ontario Study (COMP-PC): Methodology of a multifaceted cross-sectional practice-based study. *Open Medicine*.
- Davidson, E.J. (2004). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.
- Division of Health Manpower & Resources Development, Institute of Medicine. (1978). *A manpower policy for primary health care: Report of a study*. IOM Publ. 78–02. Washington, DC: National Academy of Sciences.
- Donaldson, S.I. (2003). Theory-driven evaluation in the new millennium. In S.I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 109–141). Mahwah, NJ: Lawrence Erlbaum Associates.
- Donaldson, M.S., Yordy, K.D., Lohr, K.N., & Vanselow, N.A. (Eds.). (1996). *Primary care: America's health in a new era*. Washington, DC: National Academy Press.
- Dykeman, M., MacIntosh, J., Seaman, P., & Davidson, P. (2003). Development of a program logic model to measure the processes and outcomes of a nurse-managed community health clinic. *Journal of Professional Nursing, 19*(3), 197–203.
- Forrest, C.B., Shi, L., von Schrader, A., & Ng, J. (2002). Managed care, primary care, and the patient–practitioner relationship. *Journal of General Internal Medicine, 17*(4), 270–277.
- Fraser, B., & Hollett, R.G. (2003). Use of a program logic model to guide the development of a strategic plan for Wellington County Hospitals Network. *Healthcare Management Forum, 16*(1), 12–17.
- Goertzen, J.L., Hampton, M.R., & Jeffery, B.L. (2003). Creating logic models using grounded theory: A case example demonstrating a unique approach to logic model development. *Canadian Journal of Program Evaluation, 18*(2), 115–138.

- Goldman, K.D., & Schmalz, K.J. (2006). Logic models: The picture is worth ten thousand words. *Health Promotion Practice*, 7(1), 8–12.
- Grembowski, D. (2001). *The practice of health program evaluation*. Thousand Oaks, CA: Sage.
- Haggerty, J., Burge, F., Lévesque, J.F., Gass, D., Pineault, R., Beaulieu, M.D., et al. (2007). Operational definitions of attributes of primary health care: Consensus among Canadian experts. *Annals of Family Medicine*, 5(4), 336–344.
- Health Canada. (2007). *Health care system: About primary health care*. Retrieved August 22, 2007, from <[http://www.hc-sc.gc.ca/hcs-sss/prim/about-apropos/index\\_e.html#2](http://www.hc-sc.gc.ca/hcs-sss/prim/about-apropos/index_e.html#2)>
- Hernandez, M. (2000). Using logic models and program theory to build outcome accountability. *Education and Treatment of Children*, 23(1), 24–40.
- Hogg, W., Rowan, M.S., Russell, G., Geneau, R., & Muldoon, L. (2008). Framework for primary care organizations: The importance of a structural domain. *International Journal for Quality in Health Care*, 20(5), 308–313.
- Hutchinson, B., Abelson, J., & Lavis, J. (2001). Primary care in Canada: So much innovation, so little change. *Health Affairs*, 20(3), 116–121.
- Issel, L.M. (2004). *Health program planning and evaluation: A practical, systematic approach for community health*. Toronto: Jones and Bartlett.
- Leeuw, F.L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation*, 24(1), 5–20.
- Levin-Rozalis, M. (2003). Evaluation and research: Differences and similarities. *Canadian Journal of Program Evaluation*, 18(2), 1–31.
- Mayne, J. (2004). Reporting on outcomes: Setting performance expectations and telling performance stories. *Canadian Journal of Program Evaluation*, 19(1), 31–60.
- McGilton, K.S., Fox, M.T., & Sidani, S. (2005). A theory-driven approach to evaluating a communication intervention. *Canadian Journal of Program Evaluation*, 20(2), 27–47.
- McLaughlin, J.A., & Jordan, G.B. (2004). Using logic models. In J.S. Wholey, H.P. Hatry, & K.E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 7–32). San Francisco: Jossey-Bass.

- Moyer, A., Verhovsek, H., & Wilson, V.L. (1997). Facilitating the shift to population-based public health programs: Innovation through the use of framework and logic model tools. *Canadian Journal of Public Health, 88*(2), 95–98.
- Muldoon, L.K., Hogg, W.E., & Levitt, M. (2006). Primary care (PC) and primary health care (PHC): What is the difference? (Commentary). *Canadian Journal of Public Health, 97*(5), 409–411.
- Muldoon, L., Rowan, M.S., Geneau, R., Hogg, W., & Coulson, D. (2006, Winter). Models of primary care service delivery in Ontario: Why such diversity? *Healthcare Management FORUM, 18–23*.
- Rogers, P.J., Petroschino, A., Huebner, T.A., & Hacsí, T.A. (2000). Program theory evaluation: Practices and problems. *New Directions in Evaluation, 87*, 5–13.
- Romanow, R.J. (2002). *Commission on the Future of Health Care in Canada. Building on values: The future of health care in Canada – final report* (Cat. no. CP32-85/2002E-IN). Ottawa: Government of Canada.
- Rossi, P.H., Freeman, H.E., & Lipsey, M.W. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Rowan, M.S. (2000). Using logic models in primary care reform: Navigating the evaluation. *Canadian Journal of Program Evaluation, 15*(2), 81–92.
- Rush, B., & Ogborne, A. (1991). Program logic models: Expanding their role and structure for program planning and evaluation. *Canadian Journal of Program Evaluation, 6*(2), 95–106.
- Safran, D.G., Kosinski, M., Tarlov, A.R., Rogers, W.H., Taira, D.H., Lieberman, N., et al. (1998). The Primary Care Assessment Survey: Tests of data quality and measurement performance. *Medical Care, 36*(5), 728–739.
- Schmelzle, J., Rosser, W., & Godwin, M. (2005). *Research lexicon of terms to best describe primary health care and Ontario primary health care delivery models: A summary*. Kingston, ON: Department of Family Medicine, Queen's University.
- Shah, C.P. (2003). *Public health and preventive medicine in Canada* (5th ed.). Toronto: Elsevier Canada.

- Shi, L., Starfield, B., & Xu, J. (2001). Validating the Adult Primary Care Assessment Tool. *Journal of Family Practice, 50*(2), E1.
- Starfield, B. (1998). *Primary care: Balancing health needs, services, and technology*. New York: Oxford University Press.
- Stinchcomb, J.B. (2001). Using logic modeling to focus evaluation efforts: Translating operational theories into practical measures. *Journal of Offender Rehabilitation, 33*(2), 47–65.
- Stinson, S., & Wilkinson, C. (2004). Creating a successful clinical extern program using a program planning logic model. *Journal for Nurses in Staff Development, 20*(3), 140–144.
- Stufflebeam, D.L. (2001). Evaluation models. *New Directions for Evaluation, 89*, 1–106.
- Treasury Board of Canada. (2005). *Preparing and using results-based management and accountability frameworks*. Ottawa: Government of Canada.
- Watson, D.E. (2007). *Community Health Centres in Ontario: Results-based logic model to support performance measurement, accountability and evaluation*. Toronto: Association of Ontario Health Centres.
- Watson, D.E., Broemeling, A.M., Reid, R.J., & Black, C. (2004). *A results-based logic model for primary health care: Laying an evidence-based foundation to guide performance measurement, monitoring and evaluation*. Vancouver: Centre for Health Services & Policy Research. Retrieved August 4, 2006, from <<http://www.chspr.ubc.ca>>
- Weiss, C.H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J.P. Connel, A.C. Kubisch, L.B. Schorr, & C.H. Weiss (Eds.), *New approaches to evaluating community initiatives* (pp. 65–92). Washington, DC: Aspen Institute.
- Weiss, C.H. (1997a). How can theory-based evaluation make greater headway? *Evaluation Review, 21*(4), 501–524.
- Weiss, C.H. (1997b). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation, 76*, 41–55.
- Weiss, C.H. (2004). On theory-based evaluation: Winning friends and influencing people. *Evaluation Exchange, 9*(4), 2–3. Retrieved February 17, 2006, from <<http://www.gse.harvard.edu/hfrp/eval/archives.html>>

- World Health Organization. (2005). *Global Forum IV on chronic disease prevention and control*. Geneva: Author.
- Wholey, J.S. (1979). *Evaluation: Promise and performance*. Washington, DC: Urban Institute.
- Wonca International Classification Committee. (2003). In N. Bentzen (Ed.), *Wonca dictionary of general / family practice*. Copenhagen, Denmark: Månedsskrift for praktisk lægegering.

**Dr. Margo S. Rowan** runs a consulting business and is an assistant professor in the Department of Family Medicine, University of Ottawa. She has over 25 years experience in providing evaluation and research services to health associations, governments, universities, and non-profit organizations. She has extensive experience in all facets of evaluation design, implementation, analysis, and publication.

**Dr. William Hogg** is a professor in the University of Ottawa Department of Family Medicine and Research Director of the C.T. Lamont Primary Health Care Research Center. He is a principal scientist at the University of Ottawa Institute of Population Health and affiliated scientist at the Ottawa Health Research Institute

**Lise Labrecque**, B.S.W., M.H.Sc., Cert. Prog. Eval., is a consultant in health program evaluation with 17 years experience in program development, implementation, and evaluation. She is also a health promoter at a community health centre in Ottawa, where she works with an interdisciplinary team providing front-line, primary health care.

**Dr. Elizabeth (Betsy) Kristjansson** is an associate professor in the School of Psychology, Faculty of Social Sciences, at the University of Ottawa. She is a health measurement expert whose research focuses on the spatial dimensions of health and health inequalities, food insecurity, and health and social indicators.

**Simone Dahrouge** has worked in clinical research for the past 20 years. In the last five, her work has focused on understanding the organizational structure of primary care practices and its effect on quality of care. She is currently completing the requirements for the Ph.D.