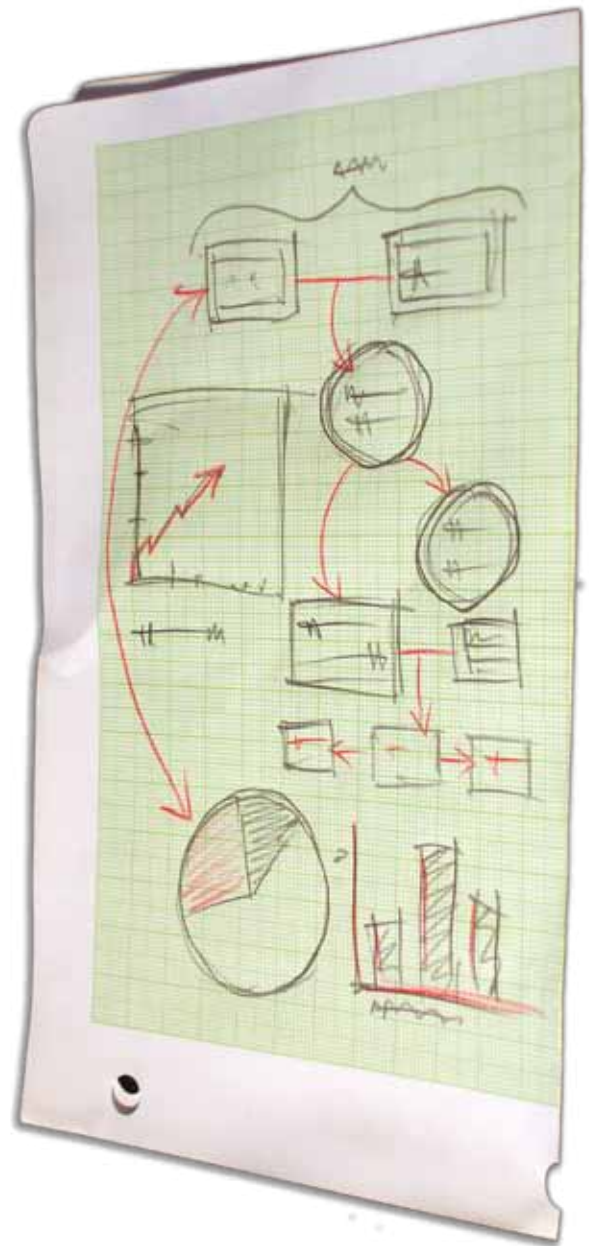# Evaluation of Healthcare Services:

Asking the Right Questions to Develop New Policy and Program-Relevant Knowledge for Decision-Making

Marcus J. Hollander, Jo Ann Miller and Helena Kadlec

Abstract

This article presents a framework for thinking about the key questions that need to be answered to develop new policy and program-relevant knowledge that can be used to make more informed decisions. It is a primer for administrators, policy makers and others about how to identify the knowledge they need to make decisions regarding new or existing programs. The article covers three related dimensions in evaluation: types of evaluations, key domains of inquiry and generic research questions. While the questions are generic, they can be readily adapted to any new and/or existing healthcare program evaluation. Examples of how the generic questions can be adapted to primary healthcare clinics and home care are presented.

rogram evaluation is an extensive topic and it is beyond the scope of this article to outline all relevant aspects. Rather, this article presents a framework for thinking about the key questions which need to be answered to develop new policy and program-relevant knowledge that can be used to make more informed decisions. Thus, this article is essentially a primer for administrators, policy makers and others about how to identify the knowledge they need to make decisions about new or existing programs. It covers three related dimensions in evaluation: types of evaluations, key domains of inquiry and generic research questions. The questions are fairly generic but, as will be shown later, can be readily adapted to the evaluation of any new or existing healthcare program.

Getting Started

Evaluation can only take place in fertile ground. This means that key actors should have an appreciation of the analysis enterprise and a desire to use analysis as a tool for decision-making. In addition, there needs to be active discussion between the evaluator and decision-makers about what new knowledge is to be developed and what research questions and operational definitions of key concepts will be used in the evaluation. (An operational definition is one which identifies objective and measurable variables as being representative of an abstract concept. For example, one may use the variable of family income as a proxy for the abstract concept of social class). In order for an operational definition to be seen as relevant and/or appropriate, key actors have to agree that it is a valid representation of the concept one is trying to measure. This can sometimes be quite contentious. Disconnects can occur if there is insufficient discussion at the outset about what new knowledge is a high priority for the administrator. There must be adequate discussion before the study is carried out, otherwise the results may be discredited because they do not answer the "real" question(s) of concern, or because there is no consensus on the validity of the operational definitions used in the evaluation. Analysis which misses the mark is disappointing for all parties.

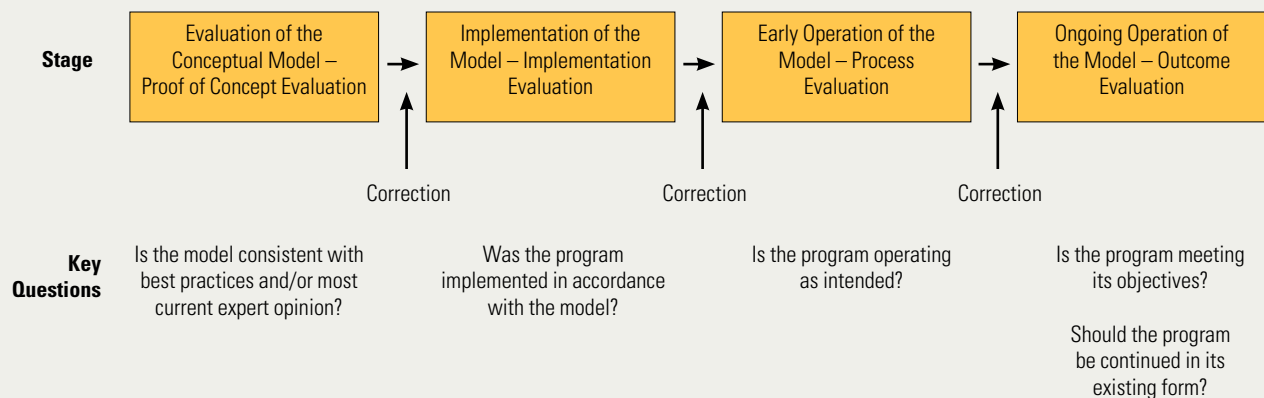## A Brief Word on Logic Models and Indicators

It is conventional wisdom that one needs to start with a logic model and indicators when conducting an evaluation. While logic models can be helpful they are not absolutely necessary as long as there is a clear description of the program and its objectives. More troubling is the trend to simplify knowledge development into creating a list of indicators as a first step in an evaluation. This approach has a high probability of producing sub-optimal results. To be useful, indicators need to be understood in a context. For example, is low cost a marker for program efficiency or a reflection of cost shifting from the organization to the client? To be relevant, indicators need to have a shared meaning, and relevance, to key actors. Unfortunately, indicators may be more a reflection of form rather than content; that is, people come up with indicators because they think they need to do so. In addition, while the indicators developed may be interesting they may not capture the essence of what administrators and policy makers need to know. A preferred approach is to develop a set of key questions which clearly reflect the new knowledge that administrators and policy makers want and need to know in order to make informed decisions. Indicators should be a product of a process of analysis. When this is the case they are much more likely to have a shared meaning and relevance to decision-makers.

Types of Evaluation

Formal evaluations of new initiatives are conducted to ensure that the initiatives are working as planned and are achieving intended results. *Process (or formative) evaluations* are conducted to determine if services are being delivered in a manner that is consistent with the model of care adopted and with the policies of the program. Process evaluations can be used to improve how services are delivered. *Outcome (or summative) evaluations* are conducted to determine if a program is meeting its stated objective(s) and/or to determine if it is better than one or more other models of care (including the model which was in place before the new program). Outcome evaluations can be used to determine the relative "worth" of a program and can be used to make decisions about whether a program will be maintained, modified or ended. Two other approaches, a *proof of concept evaluation* and an *implementation evaluation*, can also be conducted in the early stages of a new program.

Figure 1. Types of evaluation

Stage	Evaluation of the Conceptual Model – Proof of Concept Evaluation	Implementation of the Model – Implementation Evaluation	Early Operation of the Model – Process Evaluation	Ongoing Operation of the Model – Outcome Evaluation
		Correction	Correction	Correction
Key Questions	Is the model consistent with best practices and/or most current expert opinion?	Was the program implemented in accordance with the model?	Is the program operating as intended?	Is the program meeting its objectives? Should the program be continued in its existing form?

(The application of the proof of concept approach to healthcare evaluation was developed by Hollander Analytical Services Ltd. A description of this approach is presented in Appendix 1.) These evaluations should, ideally, precede both process and outcome evaluations. The first looks at the consistency of the proposed care model with best practices for similar initiatives (i.e., assesses the face validity of the model). The second evaluates the implementation phase of an initiative. It is beneficial to independently evaluate these two approaches as problems in design and/or implementation can lead to sub-optimal results.

Thus, one can think of a progression of four types of evaluation: evaluation of the model that is developed (proof of concept evaluation); evaluation of the implementation of the model (implementation evaluation); evaluation of how the model is operating (process evaluation); and evaluation of whether the model should be continued in its existing form (outcome evaluation) (see Figure 1).

The Evaluation Framework, Key Domains of Inquiry and Key Evaluation Questions

While the above provides an overview of the main types of evaluations, other approaches provide more specific domains of inquiry to be considered in conducting an evaluation. Two typologies which have been used on projects for process and outcome evaluations are those developed by the Canadian Institute for Health Information (CIHI) and the original Health Transition Fund (HTF). The CIHI performance domains are acceptability, accessibility, appropriateness, competence, continuity, effectiveness, efficiency and safety. The original HTF was a federal program to fund and evaluate new and innovative models of care delivery. As such, the HTF evaluation domains are particularly relevant for new models of care and/or pilot projects. The HTF domains are quality of service, accessibility, care coordination/integration, health impacts/effects, cost-effectiveness and transferability/generalizability. Sustainability, which refers to the extent to which a program appears to be well funded and supported and is likely to continue to exist over time, can be added to the above criteria.

We developed 11 domains of inquiry using the CIHI and HTF domains, the types of evaluation and our own experience. Table 1 presents the proposed evaluation framework and evaluation domains. Table 2 presents an initial set of generic evaluation questions for developing instruments for each domain of inquiry. The questions in Table 2 can be used to develop a series of indicators of relevance to a particular evaluation, for new or existing programs. It is recognized that not all questions may be included in each evaluation. Rather, the generic questions can serve as a guide for thinking about how any specific evaluation could be conducted. For larger programs of research, where multiple programs are being evaluated, the generic questions could be used to develop a core set of questions which should be covered in all evaluations.

Table 3 shows examples of how the framework noted above can be adapted to different types of health services.

Table 1. Types of evaluation, domains of inquiry and definitions

Types of Evaluation and Domains of Inquiry	Definition or Description
Design and Implementation of the Model	
1. Appropriateness of the model design (proof of concept evaluation and structure of the model)	This relates to whether or not the model itself is well documented, is designed to meet the stated purposes, goals and objectives of the program, and is consistent with best practices in the field. The rationale for the model, the key characteristics of the model and the organizational structure of the model are all included in this domain of inquiry.
2. Efficiency and effectiveness of model implementation (implementation evaluation)	This relates to whether or not the model was implemented in accordance with the required model design, how well or poorly the model was implemented and the acceptance of the new model by personnel in the organization and other key actors.
Functionality of the Model (Process Evaluation)	
3. Appropriate care provision	This relates to an assessment of the extent to which there are adequate staff to provide care; care provision is carried out in a consistent manner and in accordance with documented policies and procedures; and the model is "functional," that is, that the process of care provision functions in an appropriate manner.
4. Continuity of care and care coordination	This refers to how well care services, and the process of providing care, are coordinated across the component parts of the continuum.

Table 1. Continued

Types of Evaluation and Domains of Inquiry	Definition or Description
5. Competence of personnel	This relates to the professional qualifications and competence of the people managing and delivering services, for example, the care staff in a service delivery organization.
Effectiveness of the Model (Outcome Evaluation)	
6. Accessibility of service	This relates to how well, or poorly, clients can access services and/or have their questions answered. It also relates to the hours of operation and the ease of access to needed services.
7. Quality of service	This relates to the quality of the service provided in the model, and the perceptions of quality and/or level of satisfaction with the service provided by clients, family members and key stakeholders.
8. Cost-effectiveness	This relates to the value for money obtained by the organization which adopted the care model. It relates to both the costs and outcomes of the model.
9. Health impacts	This relates to the impact, if any, of the model on the clientele served and on the health status of the broader population.
10. Transferability and generalizability	This relates to the relevance of the model to other jurisdictions and/or contexts. It refers to the extent to which a given model has the potential to be adopted more broadly across Canada, and the extent to which it has actually been adopted across organizations or jurisdictions. It is a measure of the diffusion of innovation.
11. Sustainability	This relates to how well the model can continue to operate over time into the future.

Table 2. Types of evaluation, domains of inquiry and generic evaluation questions

Types of Evaluation and Domains of Inquiry	Generic Evaluation Questions
Design and Implementation of the Model	
1. Appropriateness of the model design (proof of concept evaluation)	Is the documentation on the model clear and comprehensive?
	Is the model congruent with its intended purposes and rationale?
	Is the model design congruent with the goals and objectives of the model and with best practices?
	What are the key characteristics of the model?
	What is the organizational structure of the model?
	What is the expenditure allocation, or budget breakdown, of the model?
2. Efficiency and effectiveness of model implementation (implementation evaluation)	Was the model implemented within the anticipated time frame?
	Was the program implemented consistent with the description of the model and program policy?
	During implementation were there changes to the model design? If so, were they well documented and supported?
	How well was the new model accepted by staff and management?
	Were there adjustments in the mix and/or functions of people working on the model, that is, were there human resources impacts as a result of introducing the model?
	Overall, how would the staff and management rate the "success" of the implementation?

Table 2. Continued

Types of Evaluation and Domains of Inquiry	Generic Evaluation Questions
Functionality of the Model (Process Evaluation)	
3. Appropriate care provision	To what extent is care provision consistent with program policy?
	To what extent are care needs met in a timely manner?
	To what extent are clients' questions answered in an appropriate and timely manner?
	To what extent is there adequate coverage for staff sick days and holidays?
	To what extent are emergency procedures in place and tested on a regular basis?
	To what extent are staff levels adequate to carry out the needed work?
4. Continuity of care and care coordination	To what extent is there "informational continuity" (is information from prior events used to give appropriate care to the client)?
	To what extent is there "relational continuity" (do clients/patients generally receive care from the same care provider)?
	To what extent is there "management continuity" (is care from different providers connected in a coherent way)?
	To what extent are operational reporting relationships functioning smoothly?
	To what extent does "turf protection" and/or "office politics" impact ongoing operations?
	To what extent is there effective coordination with other related organizations to the benefit of the client?
5. Competence of personnel	What percentage of staff have appropriate academic or other credentials, as required in their job descriptions?
	What is the average time key personnel have been in their positions?
	How do clients and outside stakeholders rate the competence of key personnel in the model?
	Is there adequate and ongoing orientation and training for staff working in the new model?
	Is there a regular, and ongoing, review of performance with appropriate feedback from supervisors and/or colleagues?
	Is there a continuous quality improvement program?
Effectiveness of the Model (Outcome Evaluation)	
6. Accessibility of service	To what extent is care provided over an appropriate period of time on weekdays?
	To what extent is care provided over an appropriate period of time on weekends?
	If there are waiting times for service, how reasonable are these waiting times from the client perspective?
	If there are waiting times for service, how reasonable are these waiting times from the care provider perspective?
	How consistent are waiting times with best practices for care?
	Is there evidence of ongoing efforts to reduce waiting times, as appropriate?

Table 2. Continued

Types of Evaluation and Domains of Inquiry	Generic Evaluation Questions
7. Quality of service	To what extent are clients and family members satisfied with the services they receive?
	To what extent do clients and family members perceive staff as being "caring" and "willing to go the extra mile" to met clients' care needs?
	To what extent does the program enhance clients' quality of life?
	Is there a regular accreditation of the care model or some other form of external review or evaluation regarding quality?
	Do key stakeholders perceive that the care model provides high-quality care?
	How well do staff answer questions posed by clients and family members about the condition of the client and the care services provided?
8. Cost-effectiveness	Are unit costs "reasonable" in regard to an appropriate peer group?
	To what extent is the program reaching stated (or implicit) program outcome goals?
	Has the model had an impact on human resources issues (e.g., staff reductions, retraining, etc.)?
	What are the comparative program costs and outcomes before and after the implementation of the model?
	What are the comparative costs and outcomes of the new program compared to other similar programs?
	To what extent has value for money been increased compared to the period before the program was put into place?
9. Health impacts	To what extent has the model increased the health status of those affected by the model?
	To what extent has the model increased the health status of the overall population the model is intended to serve?
	To what extent has the model identified and/or addressed key determinants of health?
	To what extent has the program reduced the rate of deterioration of the health of the clients over time (e.g., for people with chronic diseases)?
	To what extent has the program reduced the use of healthcare services (e.g., hospital admissions or readmissions, or hospital lengths of stay)?
	To what extent has the program contributed to reductions in key health indicators (e.g., infant mortality, primary care visits/costs)?
10. Transferability and generalizability	To what extent could the model be adopted in other similar settings?
	To what extent has the model served as a basis for similar developments elsewhere?
	To what extent do staff and managers believe their model will be more widely adopted?
	Which components of the model are transferable and which are unique to the context in which the existing model functions?
	What changes in structure, human resources etc. would be required in other organizations wishing to adopt this model?
	To what extent is the model bound to its context (e.g., how applicable is an urban model to rural areas, or to culturally diverse populations)?

Table 2. Continued

Types of Evaluation and Domains of Inquiry	Generic Evaluation Questions
11. Sustainability	Is this a model or pilot project with guaranteed funding only for a given period of time?
	Is ongoing funding in place in the base budget of the organization in which the model is housed?
	Are there clear and highly probable sources of alternative funding?
	What is the relative priority of the program for the larger organization of which it is a part (e.g., regional health authority [RHA]) or the funder (province or RHA which funds services directly through contracts with service provider organizations)? What would happen to the program if the broader organization, or funder, had to cut budget by 10, 20 or 30% or more?
	How well is the program supported by its clients?
	How well is the program supported by the community in which it is located and by key stakeholders (e.g., other players in the healthcare system, local leaders and politicians)?

Table 3. Examples of framework adaptations

Generic Question	Questions for a Primary Care Clinic	Key Indicators	Type of Design	Data Sources
To what extent is care provision consistent with program policy?	To what extent is multidisciplinary care provided? (This assumes a policy which mandates multidisciplinary care)	Number of staff by occupational category Ratio of FTE GPs to other professional care providers (e.g., nurses, PTs/OTs, dieticians, etc.)	Document review Analysis of administrative data	Personnel records Electronic records (to calculate FTEs)
	How frequent are breaches of policy?	Number of letters of complaint over a six-month period which indicate breaches of policy Number of breaches identified through records review	Document review Chart review	Review of letters of complaint Clinical files
Generic Question	**Question for a Home Care Provider**	**Key Indicators**	**Type of Design**	**Data Sources**
To what extent is there "relational continuity" (do clients/patients generally receive care from the same care provider)?	How well is our agency providing relational continuity?	Percentage of all visits over a three-month period provided by the person who provided the most visits Percentage of respondents who are somewhat or very satisfied with the consistency of their home care provider	Analysis of documents or electronic records Point in time, cross-sectional survey	Scheduling records Client survey

FTE = full-time equivalent; GP = general practitioner; OT = occupational therapist; PT = physiotherapist.

Conclusion

This article has presented a framework for thinking about the key questions that need to be answered to develop new policy and program-relevant knowledge that can be used to make more informed decisions. It is essentially an evaluation primer for administrators, policy makers and others and was prepared because evaluations can be done well or they can be done poorly. They can be well thought out and executed in a methodologically rigorous way or they can be reduced to an a priori list of indicators. They can add critical, new knowledge for decision-making or they can simply reflect things that are mostly already known. Ultimately, high-quality evaluations have the power to bring about significant improvements in service delivery resulting in an improved and sustainable healthcare system for all Canadians. HQ

Appendix 1. Proof of Concept Evaluation

A proof of concept evaluation is a method for externally and independently validating a new care delivery model using a panel of experts in the topic area.

Steps in Conducting a Proof of Concept Evaluation

Conducting a proof of concept evaluation involves the following steps:

1. Evaluators obtain a consensus that the parties involved in the model are prepared to participate in a proof of concept evaluation.
2. Program staff prepare a description of the model of care.
3. Using the above description, and discussions with program staff, the evaluators prepare a comprehensive yet concise description of the model so that an external expert panel can understand what the model is, why it was developed, what its component parts are and how a client would be treated in the model.
4. Evaluators conduct a literature and key informant scan to determine what best practices may exist for similar models, or what the key characteristics of good, similar models are.
5. Evaluators develop a document which outlines an "ideal," or best practices, model and its component parts based on information collected in point four above.
6. Evaluators and program staff identify key experts to be considered for the independent expert panel. Evaluators invite panel members and finalize panel membership.
7. Program staff identify a key person who does not have a vested interest in the model, but who understands both the model and the context in which it operates, to participate in the panel as a resource to the experts (e.g., a knowledgeable retired or former member of the program team with the ability to objectively answer questions about the program). This is done to ensure that the experts can clearly under-

stand the model and its context. The resource person will be required to agree to maintain confidentiality about panel discussions.
8. Convene the panel of experts for a two-day review of the model.
9. In advance of the meeting, provide an overview of the model to be assessed (see step three above) and an overview of best practices or key characteristics from other models (see step five above) to the expert panel. On the first day, facilitate discussions so that the experts develop a consensus on what characteristics or best practices would define an ideal model, that is, what would constitute a "gold standard."
10. During the second day, the actual program model is compared to the ideal template or "gold standard" and a step-by-step determination is made as to the strengths and weaknesses of each main component, or characteristic, of the model under review.
11. Based on the two-day meeting, a draft report of the findings is prepared by the evaluators and circulated back to the members of the expert panel for their review and comment.
12. Based on this input, a confidential final proof of concept evaluation report is prepared and submitted to the funder, or other appropriate group or body.

About the Authors
Marcus J. Hollander, PhD, is president of Hollander Analytical Services Ltd., a national health services and policy research company headquartered in Victoria, British Columbia. He can be reached by telephone at 250-384-2776 or by e-mail at marcus@hollanderanalytical.com.

Jo Ann Miller, PhD, is vice-president, research and evaluation, at Hollander Analytical Services Ltd. She can be reached by telephone at 250-384-2776 or by e-mail at jamiller@hollanderanalytical.com.

Helena Kadlec, PhD, is the senior scientist at Hollander Analytical Services Ltd. She can be reached by telephone at 250-384-2776 or by e-mail at helena@hollanderanalytical.com.