HTA

INITIATIVE SERIES

An Exploratory Review of Evaluations of Health Technology Assessment Agencies

AHFMR
ALBERTA HERITAGE FOUNDATION FOR MEDICAL RESEARCH
HEALTH TECHNOLOGY ASSESSMENT UNIT

Other Titles in this Series

HTA Initiative #1	Framework for Regional Health Authorities to Make Optimal Use of Health Technology Assessment
HTA Initiative #2	Making Managerial Health Care Decisions in Complex High Velocity Environments
HTA Initiative #3	Proceedings of the Conference on Evidence Based Decision Making: How to Keep Score
HTA Initiative #4	AHFMR Screening Procedure for Use When Considering the Implementation of Health Technology (Released April 2001)
HTA Initiative #5	Priority Setting in Health Care: From Research to Practice
HTA Initiative #6	AHFMR Screening Procedure for Use When Considering the Implementation of Health Technology (Released April 2002)
HTA Initiative #7	Local Health Technology Assessment: A Guide for Health Authorities
HTA Initiative #8	Minimally Invasive Hip Arthroplasty
HTA Initiative #9	Elements of Effectiveness for Health Technology Assessment Programs
HTA Initiative #10	Emergency Department Fast-track System
HTA Initiative #11	Decision-Making for Health Care Systems: A Legal Perspective
HTA Initiative #12	Review of Health Technology Assessment Skills Development Program
HTA Initiative #13	Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields
HTA Initiative #14	Workshop Summary Knowledge-Brokers: Linking Researchers and Policy Makers
HTA Initiative #15	Quantitative Approaches to Patient Safety. Research in Risk Analysis and Risk Management as Applied to Radiotherapy

For more information contact:

Health Technology Assessment Unit
Alberta Heritage Foundation for Medical Research

Suite 1500
10104 – 103 Avenue
Edmonton, Alberta
Canada T5J 4A7

Tel: 780 423-5727
Fax: 780 429-3509

HTA Initiative #16 ● February 2006

An Exploratory Review of Evaluations of Health Technology Assessment Agencies

Margaret Wanke
Don Juzwishin
Richard Thornley
Liza Chan

AHFMR
ALBERTA HERITAGE FOUNDATION
FOR MEDICAL RESEARCH

HTA Initiative #16
An Exploratory Review of Evaluations of Health Technology Assessment Agencies

Prepared by:
Margaret Wanke, Don Juzwishin, Richard Thornley, Liza Chan

Acknowledgements

The Alberta Heritage Foundation for Medical Research is most grateful to the following persons for review and provision of information and comments on the draft report. The views expressed in the final report are those of the authors:

Dr. Ansgar Gerhardus
Institut für Gesundheit-und Medizinrecht, Germany

Dr. David Hailey
Senior Advisor, HTA, AHFMR
Australia

Professor Arnold Love, PhD
Children's Mental Health Services
Toronto, ON

Dr. Berit Morland
Norwegian Knowledge Centre for the Health Services, Norway

The authors would like to thank the INAHTA members who took the time to respond to our survey and review our tabulations to ensure that they were accurate.

ISBN 1-894927-30-3 (Print)
ISBN 1-894927-31-1 (On-Line)

ISSN: 1706-7855

Additional information and comments relative to the initiative report are welcome and should be sent to:

Director, Health Technology Assessment
Alberta Heritage Foundation for Medical Research
1500, 10104 – 103 Avenue NW
Edmonton, Alberta, Canada T5J 4A7

Tel: 780-423-5727
Fax: 780-429-3509

www.ahfmr.ab.ca

Alberta's health technology assessment program has been established under the Health Research Collaboration Agreement between the Alberta Heritage Foundation for Medical Research and Alberta Health and Wellness.

TABLE OF CONTENTS

Tables and Figures

INTRODUCTION

In a time when the accountability for public expenditures is at a zenith and public funders are evaluating the effectiveness or impact of the programs and services they fund – it was thought opportune to examine evaluations conducted of HTA agencies with a view to informing why and how the evaluations were conducted and what changes resulted from the evaluations.

To date no comprehensive systematic review or analysis has been conducted of the evaluations done of health technology assessment (HTA) agencies around the world. We expect that evaluations or reviews of HTA agencies are conducted for many different reasons. HTA is a reasonably young endeavour and reviews are usually used as a continuous improvement or accountability initiative to identify ways to improve practice or to determine whether the funding agency is receiving value from the expenditure. The increased rapidity of innovation and diffusion of technologies in contemporary societies is, on the one hand, increasing the demand for health technology assessments and, on the other hand, creating a need to demonstrate that the outcomes or impact of HTA products are relevant to policy and decision makers, as well as practitioners.

The authors of this study collected and reviewed the external evaluations conducted of members of the International Network of Agencies for Technology Assessment (INAHTA) between the years 1994 to 2004. In April 2005 there were 39 INAHTA member agencies in 21 countries.

The objectives of this review were:

1. To propose a generic evaluation framework for HTA agencies to strengthen their evaluation capacity.
2. To conduct a review of HTA agencies to understand what aspects of HTA agencies have been evaluated, approaches/methods used, outcomes of the evaluations and to understand what was learned through these evaluations that could guide HTA agencies to best serve their mandate.

ORGANIZATION OF REPORT

The subsequent chapters of this report are organized as follows.

In the section *Selected Literature Review Findings*, the results of our literature search for potential HTA agency evaluations and conceptual frameworks are reported.

In the *Conceptual Framework* section, we describe our final overall conceptual framework.

In the section *Review Findings*, the results of the review are presented, organized by the elements contained in the overall conceptual framework.

In the *Discussion* section, we present the limitations of the study and our general observations of the study findings.

The *Reflections* section contains a summary of our observations and suggestions according to the study objectives.

SELECTED LITERATURE REVIEW FINDINGS

Conceptual framework

The authors sought to identify a generic framework or model potentially applicable to the evaluation of programs that could be instructive in the evaluation of HTA agencies. Three useful frameworks were identified.

In 2001, Stufflebeam proposed a classification of 22 evaluation approaches.[1] He defines program evaluation as "a study designed and conducted to assist some audience to assess an object's merit or worth" (p. 11) and notes that while such a definition is congruent with dictionary definitions and professional standards in the field of evaluation, many studies undertaken in the name of program evaluation do not necessarily conform to this definition.

Stufflebeam's 22 categories are grouped under four major headings. The first heading – *pseudo-evaluations* - represent those with incomplete or invalid findings, often undertaken for public relations or political objectives.

The second major category of approaches – *question and methods-oriented* – generally address specified questions or employ specific methods. In this category, Stufflebeam includes objectives-based, accountability (such as payment by results), experimental, case study, benefit-cost studies as well as program theory based evaluations.

Of note in this second category are two forms of evaluation not widely reported in the program evaluation literature – clarification hearings, and criticism and connoisseurship. Clarification hearings adopt a judicial approach where a program is placed on trial, with arguments for and against the program presented by role-playing evaluators who have compiled their evidence in advance of the trial. The jury is generally represented by the program's stakeholders. The criticism and connoisseurship approach, noted by Stufflebeam to have evolved from art and literary criticism, uses experts in the evaluation's subject area to undertake an in-depth analysis that involves describing, critically appraising and illuminating a particular program's merits, salient features and problem areas (p. 36).

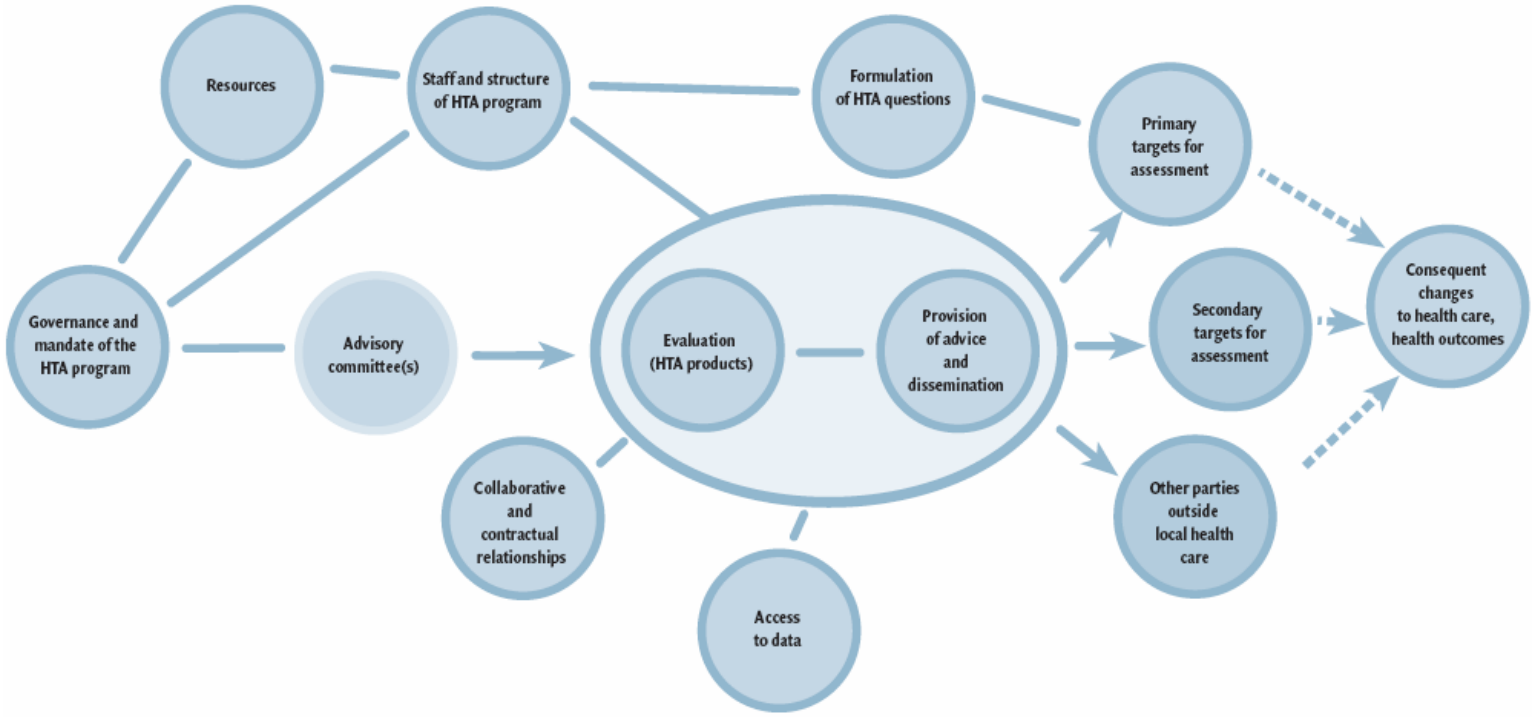
The third category of evaluation approaches – *improvement/accountability-oriented* – stress comprehensiveness in the evaluation of a program's merit and worth. Employing a full set of questions, the evaluators "seek to examine the full range of pertinent technical and economic criteria for judging program plans and operations. They look for all relevant outcomes, not just those keyed to program objectives", (p. 42) and may look to the assessed needs of the program's stakeholders as criteria for judging its merit or worth. Generally, the evaluators have an objectivist orientation "and assume an underlying reality in seeking definitive, unequivocal answers to the evaluation questions" (p. 42). Often they use mixed quantitative and qualitative methods.

Stufflebeam includes decision/accountability-oriented studies, consumer-oriented studies and accreditation/certification approaches in this category.

Finally, Stufflebeam's fourth category of evaluation approaches – *social agenda/advocacy approaches* – are those evaluations that are focused on making a difference to society. An underlying objective to the program evaluation is ensuring that "all segments of society have equal access to educational and social opportunities and services. They have an affirmative action bent toward giving preferential treatment through program evaluation to the disadvantaged" (p. 62). Stakeholders are usually heavily involved in the design, collection and interpretation of findings. Unlike the previous category, these evaluation approaches "eschew the possibility of finding right or best answers and reflect the philosophy of postmodernism, with its attendant stress on cultural pluralism, moral relativity and multiple realities" (p. 62). Examples of these approaches are client-centred studies/responsive, constructivist, deliberative democratic and utilization-focused evaluation.

The second model influencing this study's conceptual framework came from the HTA literature. Hailey explored the question of the effectiveness of HTA agencies, stating that "together … HTA products and their dissemination – form the primary areas for determining effectiveness of an HTA agency" (p. 4).[2] Drawing from the HTA and organizational theory literature, Hailey moves beyond the products and their dissemination to develop a model which includes governance, staff and structure, resources, advisory committees, collaborative and contractual relationships, access to data, formulation of HTA questions, primary targets for assessment, secondary targets for assessment other parties outside local health care and consequent changes to health care as determinants of a HTA program's effectiveness (see Figure 1). To make the model useful, Hailey operationalized the determinants of effectiveness into approaches that those evaluating an HTA program or agency could use (pp.18-26).

Figure 1: Determinants of effectiveness



A third type of model drawn on for the development of the conceptual framework was that of program logic models reported in the program evaluation literature. Logic models can be used to clarify the purpose and assumptions of a program, schematically represent the linkages and depict the underlying rationale or logic of a program.[3] A program's "logic" is presented by specifying the resources, inputs or structures that are used to undertake the program's processes or activities. In turn, these processes result in the development of products or program "outputs". The outputs are expected to result in specific outcomes, which may be classified in terms of their immediate (short-term), intermediate or ultimate (long-term) impacts.

In addition to the general frameworks or models, considerable discussion of one aspect of the evaluation of HTA agencies – impact assessment – was evident in the literature. Hailey, Cowley, and Dankiw[4] and Drummond[5] were the first to explore and identify factors for increasing the impact of HTA.

Hailey, Cowley, and Dankiw[4] argued that while health technologies were receiving increasing attention through systematic assessment, little was known about the effectiveness of HTAs in terms of changes to health care planning or the diffusion of technologies. They identified considerations and issues associated with such impact assessment, including the fact that HTA is not the only factor influencing policy decisions and the long-term nature of some of the effects of HTA, such as changes in expectations and behaviour patterns of users and providers. The authors identified a range of impacts of a technology from denial of access through controlled introduction, effect on range of application and, finally, encouragement of appropriate technology usage. They identified possible measures for each type of impact along with political conditions applicable to the assessment of the impact. For example, the impact "denial of access to technology" could be measured by the subsequent presence or absence of the technology. However, when considering such assessment, one must consider factors such as legislation, political will and possibly professional support.

Drummond[5] identified the following to be taken into account for assessing the impact of an HTA agency its service and products:

- timeliness
- validity
- involvement of decision makers
- suitable dissemination
- availability of policy instruments
- recognition of the existence of complex incentives (p. 78).

Jacob and McGregor[6] approached the issue of evaluating impact of health technology assessments with three questions:

- What impact was intended?
- To whom was the message directed?
- To what extent was the hoped for impact achieved, first in terms of policy and second in terms of actual distribution and utilization of the technology. (p. 69)

The authors go on to point out some of the difficulties of determining impact; for example, attribution of cause and effect, multi-factorial confounding influences, and time lapse.

Lehoux, Battista, and Lance[7] recognized the need for ongoing monitoring by HTA agencies, stating "that the normative function of HTA and its impact on society and the health care system must be more closely examined" (p. 24). This is because HTAs influence depends on how the users and other actors react to an assessment, agree with or contest its content and subsequently negotiate policy solutions. Their approach combines elements of self-assessment and external impact assessment in a comprehensive approach that takes into account the multiple players and perspectives inherent in the HTA sociopolitical context.

In his masters thesis on the topic of assessing the impact of HTA reports, Gerhardus identifies types of impact, indicators, and methods for extraction in three approaches or models to health technology assessment: rational model, process model and enlightenment model.[8] By cross tabulating the first three categories against the three approaches, Gerhardus derives a toolkit for evaluating impact of health technology assessments.

In keeping with a commitment toward continuous improvement in technology methods, a framework for reporting on HTA impact has been developed by the International Network of Agencies in Health Technology (INAHTA). As part of a pilot project initiated in 2004, INAHTA members were asked to provide information on HTAs via the members' only section of the INAHTA web site. They were asked to complete an HTA impact form not less than six months after the HTA report's publication date. The form asks about a number of impact indicators, relating to government decisions at the regional, national or international level. Positive, interim and negative indications may also be reported. As of April 2005, the impact of 69 reports were documented. Experience with this initiative was reviewed at the 2004 Annual Meeting. The impact form is reproduced with permission in Appendix C.

Evaluations of HTA agencies

Our literature review identified four evaluations of HTA programs or agencies. Battista, Feeny and Hodge describe their evaluation of the Canadian Coordinating Office for Health Technology Assessment (CCOHTA) conducted in 1993.[9] The authors point out that "conceptually, methodological standards much like those used to evaluate diagnostic technologies would be applicable to evaluating a health technology assessment agency" (p. 103). The methodological, time and resourcing challenges of identifying the improvement in patient outcomes or the health of the population in undertaking such an evaluation are noted and, given the challenges, the authors opt to focus their evaluation on process rather than outcomes. In spite of this limitation, the authors hope that "results reported here should be viewed as evaluating the evaluators as a means to evaluating a particular technology assessment agency. Sharing this evaluation with the technology assessment community is a hopeful step toward enhancing evaluation of both process and outcomes in technology assessment" (p. 103).

In a review of the first four years of the Quebec Council on Health Care Technology (CETS), Jacob and Battista reported on an evaluation that looked at whether the organization performed its mandated functions and attempted to measure the organization's impact on decision making.[10] The evaluation reported that CETS performed well and also saved the health system approximately $25 million from their advice, half in recurring annual costs. Several limitations of the study are noted, "first, few assessment criteria have been established in the relatively new field of technology assessment" and that "enough time may not have elapsed for their full effect to be felt (p. 570). The authors warn "reconciling and preserving scientific and political credibility will be a permanent challenge for health care technology assessment" (p. 571).

These two evaluations took place prior to 1994 and were, thus, excluded from our review.

Two additional evaluations met the selection criteria and data from these articles were included in our analysis. Dixon and colleagues set out to examine the influence of the English National Health Service South and West region Development and Evaluation Committee (DEC) technology appraisal reports on purchasing and clinical decisions.[11] The authors concluded that health service staff perceived that the DEC process had an impact on policy decisions and clinical practice but actual impact on practice could not be ascertained with the available data (p. 18).

Jacob and McGregor published the results of a second evaluation of the CETS that was completed in 1995.[6] The authors, building on the work of Jacob and Battista, turned the focus of their study toward developing a methodology that evaluated the impact of health technology assessments by looking at 21 HTAs and their impact on policies and diffusion. Jacob and McGregor note that "however excellent an HTA may be, if it fails

to influence the working of the health care system, it is without impact and must be considered without value" (p. 69). The authors note limitations of evaluating impact on policy, such as, identification of critical incidents, categorization of health care policies, and the systematic use of documentation which facilitated a degree of objectivity on one hand but resulted in limitations in relying on analysts' judgment on the other. Triangulating with interviews with key actors helped to establish objectivity. Another limitation encountered is that of attribution as HTAs may not be the only source of influence. The authors conclude the paper by stating "the best insurance for impact is a request by a decider that an evaluation is made" (p. 78), that is, if a policy or decision maker requests a specific HTA, chances are they will use it.

Two articles describing the content analysis of HTA documents across agencies in Canada were identified but not included in our review as their focus was not on individual agencies.[12,13]

CONCEPTUAL FRAMEWORK

Figure 2 depicts a framework that represents the design, product and implementation considerations for HTA agencies as they contemplate past, present and future evaluations, based on elements found in the literature and further identified through the review process. Incorporating many of the 22 approaches proposed by Stufflebeam, we have proposed a multidimensional classification, based on our experience with classification of the 16 HTA evaluations.

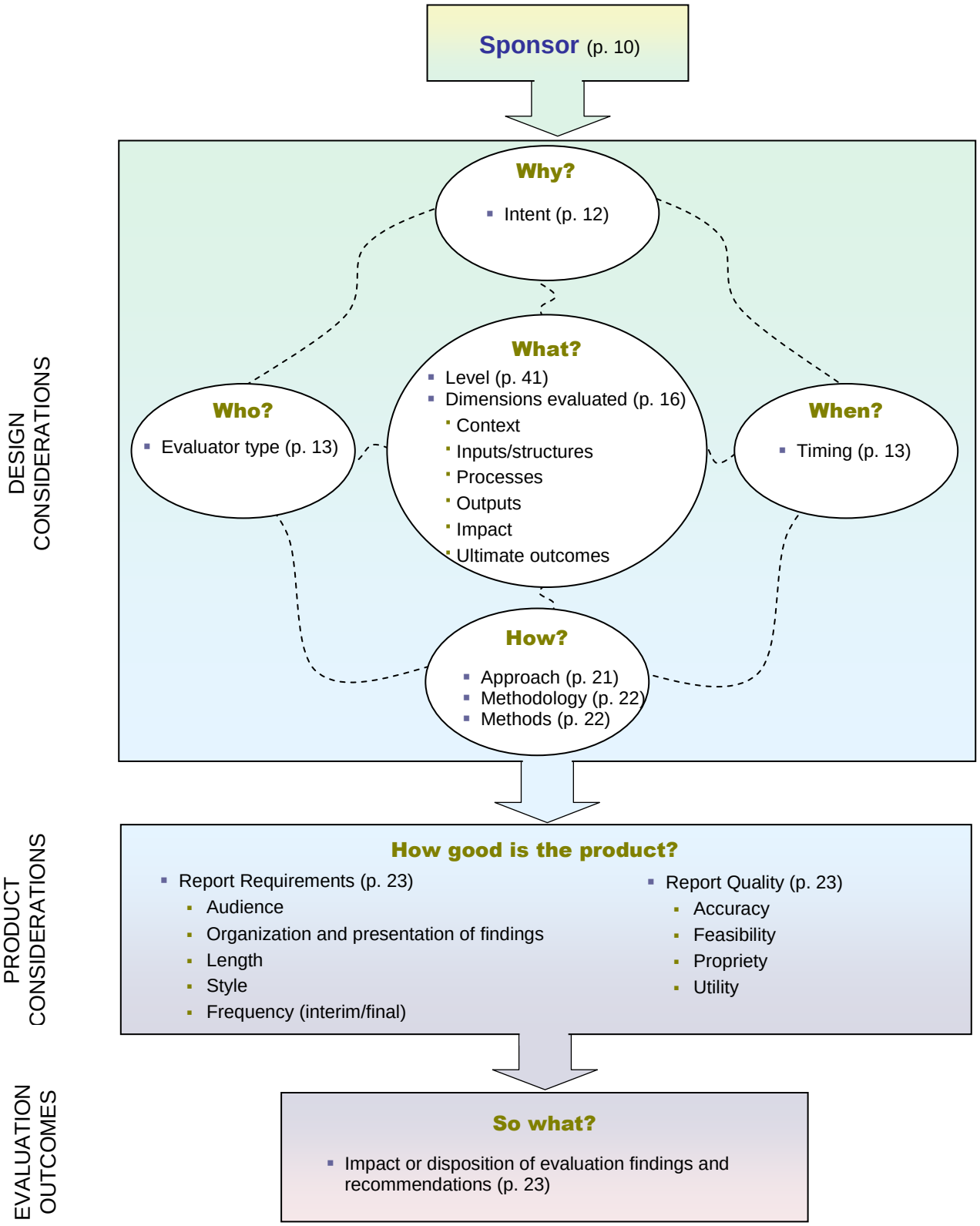
The framework includes three main categories: design considerations that guide the development of the evaluation approach and plan; product considerations that sponsors can use to specify reporting requirements and evaluate report quality; and finally, a feedback loop that considers the outcomes or disposition of the evaluation findings and recommendations. Each dimension of the framework is discussed in turn. This framework is also used to present the results of the review in the next chapter.

Sponsor

Sponsors of evaluations may represent internal, external or joint internal and external audiences.

- The internal sponsor, in the case of HTA agencies, would be the agency itself or the organization in which the HTA agency resides.
- Sponsors that are external to the program or organization may include funders (in this case, generally government funders), other stakeholders with a direct interest in the agency, academic institutions or research organizations, independent researchers or evaluators, or societal representatives (e.g., a parliamentary committee or citizen group).
- Joint internal and external sponsorship may be undertaken when there is a mutual interest or stake in the results of an evaluation.

Figure 2: Conceptual framework



Design considerations

Intent

Intent is the first design consideration. This element refers to the underlying driver or often unstated purpose of the evaluation. This aspect is critical as it can and should influence the other evaluation considerations – who, what, when and how.

Table 1 outlines six possible underlying rationale or driving forces for conducting an evaluation (adapted from Stufflebeam[1]). It is suggested that sponsors and evaluators be clear at the outset about the fundamental intent of the evaluation.

Table 1: Possible evaluation intents

Intent/Agenda	Main Objective	Characteristics or Examples
Accountability	Account for ”investment” in agency Determine if pre-determined objectives or requirements are met	• Stress the need to assess a program’s merit or worth, sometimes related to a decision for continuation or discontinuation • Employs external, independent perspective • A primary concern may be whether a funded agency is meeting its contractual obligations, and the assessment may be related to a decision regarding continuation • Involves mostly a retrospective orientation • Many may include pass/fail standards (or minimum standards of performance), payment for good results or sanctions for unacceptable performance.
Political agenda/ public relations	Convince constituents about the value or merit of the program/agency	• Lack of full disclosure of findings • Biased instruments, processes or interpretations • Manipulative use of data
Problem-solving/ Issue resolution	Address specific questions	• Issues driven – public or sponsor concerns may have stimulated the need for a closer look at the agency • Narrowly defined questions
Quality improvement	Strengthen agency	• Retrospective with view to future (e.g., needs assessment or strategic planning approaches) • May be more likely than accountability approaches to employ internal evaluation resources.
Research/curiosity	Generate knowledge	• General purpose is to add to the body of knowledge without drawing specific conclusions or recommendations about the value or merit of the agency
Social agenda/ advocacy	Improve or make a difference in society	• Concerned with societal issues, such as inequity • Bent towards reform or affirmative action • Involvement of stakeholders in design, collection and interpretation of data • Includes consumer-driven, constructivist and deliberative democratic evaluation approaches • Will sacrifice objectivity in favour of a democratic process, recognizing the many realities of a pluralistic society and multiple realities

Timing

This element of the model involves three dimensions: timing and length of the evaluation within a calendar or fiscal year; timing given financial, organizational and political factors; and type of evaluation undertaken at the particular time in the life cycle of the agency. The first consideration is largely logistical, relating to factors such as the scope of the evaluation and resource availability.

The second consideration relates to contextual factors that highlight the need, opportunity or readiness for an evaluation. For example, evaluations undertaken for the purposes of accountability may be timed according to a pre-determined schedule established under a contractual agreement.

The final consideration is the natural cycle relating to the type of evaluation undertaken, given a program or agency's phase in its lifecycle. Formative evaluations are often undertaken in the first years of development with an objective of ensuring all processes, inputs and structures are functioning at an optimal and stable level before a summative or outcomes evaluation is attempted. An implementation evaluation may be undertaken before or as part of a formative evaluation to determine whether the program has been implemented as conceptualized and planned.

Evaluator type

Evaluators are generally selected based on their expertise and skill set in any of the following areas, which are not exhaustive or mutually exclusive:

- ***Content expert*** in the subject area of the evaluation (in this case HTA).

  This category is consistent with Stufflebeam's *criticism and connoisseurship* approach whereby "certain experts in a given substantive area are capable of in-depth analysis and evaluation that could not be done in other ways" (p. 36).[1] These types of evaluators will use their "perceptual sensitivities, past experiences, refined insights and abilities to communicate their assessments" (p. 36). Stufflebeam notes the main advantage of this type of approach is the level of detail and insight that can be brought forward to the sponsor. The primary disadvantage noted is that the study relies heavily on the qualifications and expertise of the particular content expert "leaving room for much subjectivity" (p. 37).

- ***Program evaluators:*** generally hired or contracted by sponsors because of their expertise in the field of program evaluation.

  A program evaluator's orientation will be to design an evaluation to meet the expectations of the client. Many will use a program theory-based or logic model approach. As a general rule, program evaluators will be less concerned with publishing the results of the evaluation than that their recommendations are acted upon by the sponsor or agency. The advantage

of using experienced program evaluators is their knowledge of evaluation methods and methodologies, objectivity and rigour. Because they are generally not content experts, however, they may not be able to offer the level of insight into the program or agency's issues as might a content expert.

- ***Researchers:*** often affiliated with academic institutions or research organizations.

  Typically, these types of evaluators will approach an evaluation as a research study, and will likely have an interest in publishing the results of the study. They may, through a research grant, themselves be the evaluation sponsor. Generally, researchers are well qualified for scientific study and undertake comprehensive and rigourous evaluations. This comprehensiveness and rigour may result in longer timelines for an evaluation than might be evident in the other studies. In some cases, the researchers may be driven at least as much by the desire to publish as to improve the program or agency's performance.

- ***Methodologists***: individuals who specialize in particular methodologies.

  Examples of methodologists include health care economists, data analysts and psychometricians. Generally, these types of evaluators possess a specialized methods-based skill set. A particular specialist may focus on a single or narrow range of evaluation questions, or apply the same methodologies across different studies. They may play a supportive role in a more comprehensive, broadly focused evaluation.

- ***Project managers***: often used for internal evaluations and certain types of evaluations, such as those based on management information systems.

In some cases, evaluators may not fall into a particular category but be selected because of the type of evaluation approach chosen. This is true, for example, in a judicial approach used in one of the evaluations reviewed where the evaluators represented a parliamentary committee.

Level at which evaluation is directed

Evaluations may be directed at one or more levels, as follows:

- ***System***

  This represents a collection of independent but inter-related organizations, programs and/or resources that together, comprise a unified whole. Due to the complexities involved, the number of such evaluations reported is relatively few when compared with evaluations at other levels. One example of a framework for such an evaluation is Cumper's *Evaluation of National Health Systems* (1991).[14]

- ***Network/Partnerships***

  Our increasingly complex and information-driven era has enabled rapid, effective communication across organizational and geographic boundaries. One result of this phenomenon is an ever-increasing number of networks, partnerships or clusters. Networks and partnerships may be formal, governed by a legal contract or an informal network of agencies, programs or stakeholders with a common interest, who agree to abide by a common set of principles. The International Network of Agencies for Health Technology Assessment (INAHTA) and national Cochrane Networks are examples of networks of organizations that could be evaluated.

- ***Organization***

  Organization-oriented evaluations focus on a particular agency or organization. Often these are called organizational reviews, and in the business environment, may focus on the efficiency and effectiveness of business processes, customer satisfaction, and whether the organization's strategic and financial targets are met. Kaplan and Norton's *Balanced Scorecard* approach is an example of a framework for such an evaluation.[15]

- ***Program***

  A program is a system of projects or services with defined resources and processes targeted to achieving specified objectives. A program usually resides within a larger organization. This is the level at which most evaluations are reported in the literature.

- ***Sub-program***

  Frequently, a program may be divided into a number of meaningful sub-units of activities. If this is the case, an evaluation may focus on only one of the sub-units. For example, if an HTA programs chooses to conduct field evaluation or primary research, this may be organized as a distinct sub-program alongside the health technology assessment sub-program.

- ***Individual***

  Assessment of individual staff or student performance, such as achieved through performance appraisals and student examinations are a form of evaluation, and when aggregated, may be used in an evaluation of a program or organization (e.g., school achievement examinations).

In this review, the level of evaluation under study could be described as the program or organization, depending on where the HTA agency sits within its governing body. The reviewers used a logic model approach to depict the dimensions of an agency being investigated. While program logic models are generally associated with the program

level, they may also be applied at the other levels of evaluation. Logic models remain one of the most flexible vehicles for describing the unit under evaluation and selecting the priority aspects to be evaluated.

Because of the interest in the sustainability of HTA agencies as an underlying query for this review, the reviewers documented the extent to which evaluations addressed the relationship between the HTA agency and the broader system level represented by the technology diffusion cycle (see Figure 3).

Dimensions of evaluation of HTA agencies

Figure 3 depicts the range of dimensions conceived for an evaluation of an HTA agency.

The Level 1 model depicts the context in which HTA agencies/program operate. This diagram, adapted from Menon and Stafinski, illustrates the diffusion process from technology innovation through research and development, assessment, policy or practice development and implementation, and finally to obsolescence or replacement.[16] Generally, the elements identified in the Level 1 model would not be the focus of an HTA agency or program evaluation, as most of these are beyond the direct control of the agency. However, because socio-political factors play such an important role with respect to HTA, evaluators of HTA agencies, particularly of impact, will need to consider contextual factors influencing the operation and success of the HTA agency/program.

In the Level 2 model, the HTA function which operates within the Level 1 technology diffusion process is depicted as a logic model, including inputs/structures, processes, outputs, outcomes/impact and ultimate impact. The items listed under these headings represent the menu of items upon which an evaluation may be focused. A comprehensive evaluation may touch on most of the elements listed. A more narrowly focused evaluation may seek in-depth understanding of one aspect of the program or agency, such as may be the case in an impact assessment.

Generally, evaluation undertaken in the early years of implementation should focus on aspects of the program's inputs/structures and processes. Once these are fully established, a program may logically begin to shift its evaluation focus onto the program's outputs. Finally, when a program has matured and is stable, evaluation of impact should be undertaken with a view to determining the program's merit or worth.

Inputs/structures

Inputs may be defined as

> *"Resources (human, material, financial, etc.) used to carry out activities, produce outputs and/or accomplish results."*[17]

Structures may be defined as the mechanisms used to organize and account for the activities. Types of structures include accountability, governance, organizational and committee.

Column 1 of the Level 2 model identifies the inputs and structures of HTA agencies. These include the agency's mandate, the principles and values under which it operates (e.g., transparency and independence), governance and accountability structures, contractual relationships with funders and suppliers, collaborative relationships with stakeholders external to the agency/program (e.g., economists, ethicists, engineers, public), financial and human resources, committee and board structures, datasets or information systems, and sources of requests.

Processes

Processes may be defined as:

> *"An operation or work process internal to an organization, intended to produce specific outputs (e.g., products or services). [Processes] are the primary link in the chain through which outcomes are achieved."*[17]

The processes that consume or are influenced by inputs and structures are divided into two categories: generic management processes and processes specific to HTA. The latter include the selection and prioritization of HTA topics, question formulation, commissioning and monitoring, data collection and analysis, formulation of guidance or recommendations, report preparation and review, dissemination, and appeals.

Outputs

Outputs are:

> *"Direct products or services stemming from the activities of a policy, program or initiative, and delivered to a target group or population."*[17]

Traditionally, HTA agency processes have been directed towards HTA products represented by full or time limited assessment products. More recently, other HTA products have emerged as agency outputs, including quick response reports (e.g., list of systematic reviews undertaken on topic of interest without assessment) and horizon scans. Outputs may be described and evaluated in terms of topic (including relevance), quantity, quality (including accuracy, comprehensiveness, format, consistency in structure, timeliness and relevance) and cost. Outputs may also include products not generally associated with HTA agencies, such as student placements completed and number of workshops presented with an objective of building HTA capacity. Under this category of evaluation, impact would not be considered.

Impact

Impact may be conceptualized as a series of steps in a chain of results. At the first level of impact, stakeholders are aware of the existence of the agency or products. Awareness may lead to an attitude change regarding HTA, ideally representing acceptance of the agency or HTA products. The final sub-category of the first level is satisfaction with the agency or products. None of the sub-categories at the first level suggests actual application or use of the products or findings.

The second level of impact is represented by evidence of actual use or application of the HTA products. Lavis describes three types of utilization: symbolic, conceptual and instrumental.[18] Symbolic utilization applies when evidence is used to serve political purposes. In this case, evidence may be found and strategically used to support a particular policy position. Conceptual utilization represents a change in awareness, knowledge or attitude about the technology. In an HTA example, a health service provider may identify that particular HTA reports has provided new knowledge about the effectiveness or safety of a particular technology. Finally, instrumental utilization represents a change in policy or practice due to the HTA product. In the authors' experience it is this last form of utilization that HTA researchers often signify when they refer to a product's "impact".

In the third level, impact of the HTA product could be measured beyond its immediate objective of influencing policy or practice, to a broader influence on the technology itself. This involves impact on the research and development, adaptation or obsolescence of a technology.

The impact (4th) column of Level 2 is partially shaded to address the issue of attribution. HTA programs and agencies can be only partially responsible for the acceptance, utilization and impact of its products. The socio-political environment in which HTA agencies operate include a multitude of perspectives, interests and agendas. It is possible that despite having state-of-the-art resources, processes and products, the political climate is not conducive to the acceptance and utilization of an HTA product. The U.S. OTA example offers a prime example. Various authors providing a retrospective on the dissolution of this agency suggested the agency's demise related more to the particular agenda of the governing party than to the success or failure of the organization.[19,20] Evaluators need to consider the extent to which impact can be attributed directly to the HTA agency or program.

Ultimate outcomes

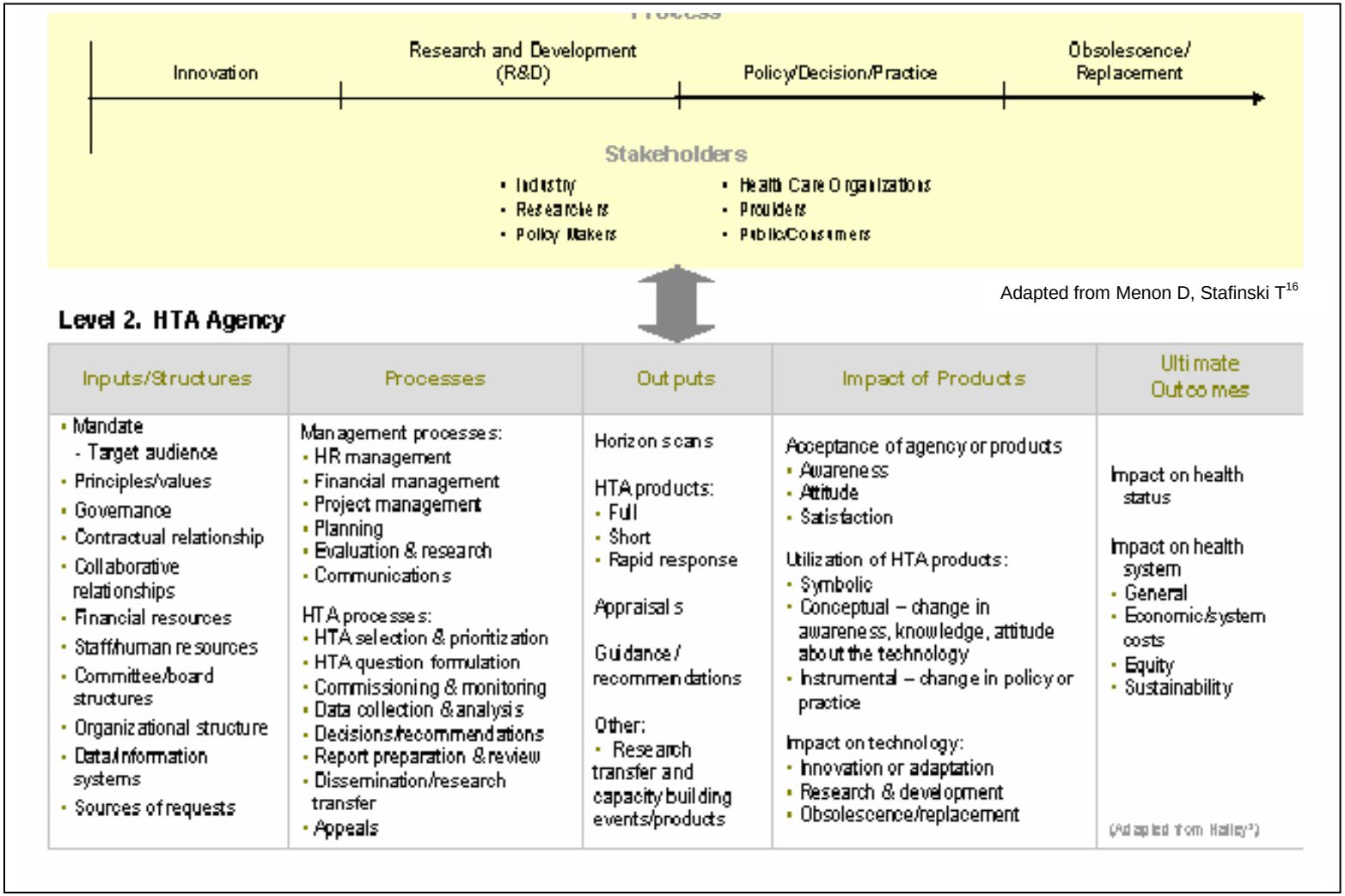
Evaluation at this level looks beyond the impact of individual HTA products, and explores the impact of the HTA agency or mandate, generally, on health of the population and the sustainability of the health system. This form of evaluation generally takes a societal perspective, as described in the subsequent section on Approaches.

Ultimately, HTA agencies/programs strive to influence the health status of individuals or a population through the availability of effective, safe and cost-effective technologies. In most countries or jurisdictions, the HTA agency or program also strives to influence the health system in terms of economic impact, equity and/or sustainability.

Because the factors influencing population health status and the health system are numerous, outcomes at these levels cannot, as a general rule, be directly attributed to

HTA agencies or programs. This is represented in the diagram by the lack of shading of this column.

Figure 3: Dimensions of HTA agency/program evaluation



Approach

Eight approaches to an evaluation, adapted from Stufflebeam[1] and Kaplan and Norton[15] based on our analysis of the review findings are described in Table 2.

The overall approach frequently reflects the orientation or inherent perspective the sponsor and/or evaluator brings to the evaluation. While there is generally a close relationship between the type of evaluator and approach, this is not always the case. For example, it is possible that a researcher may adopt a program evaluation approach, or a program evaluator adopts a business approach when evaluating organizations rather than programs.

Table 2: Approach

Approach	Characteristics
Research	▪ Experimental or quasi-experimental design, and may be presented as a research study. ▪ May be methods-oriented (e.g., economic analyses). ▪ May be published in the formal literature.
Business	▪ Evaluator is concerned with the performance of an organization, and may incorporate financial, consumer and internal business operations perspectives. ▪ Examples include organizational reviews and balanced scorecard approaches.
Constructive critic	▪ The program is judged based on an analysis undertaken by someone with recognized expertise in the subject matter. ▪ Stakeholders do not necessarily desire an objective and impartial evaluation; rather they may explicitly desire the informed opinions of a credible expert or experts.
Education/learning	▪ Performance testing of individuals where knowledge gained is used to measure effectiveness of a program or agency. Examples include student examination scores and pre-post knowledge questionnaires related to educational or training programs.
Informatics	▪ Evaluation is driven by available data on management information systems (i.e., what evaluative conclusions can be drawn from this dataset?
Judicial	▪ Program is put on trial, and arguments for and against are presented to a "judge" or "jury" which issues a ruling on the program's success or failure.
Program evaluation	▪ Generally, this approach follows guidelines for evaluation as specified by national or international professional evaluation organizations. ▪ May be based on program objectives or on questions presented by the evaluation sponsor. ▪ May be based on program theory or model, such is noted in the use of logic models.
Societal	▪ The evaluator is or represents the consumer or society's interest. ▪ May be consumer driven, use a deliberate/explicit democratic framework or otherwise demonstrate concern for the best interest of society as a whole. ▪ Considers and incorporates broad range of stakeholders in the evaluation process, and is inclusive rather than exclusive. ▪ Holistic, environmental, organic rather than narrow, internally focused and prescriptive. ▪ Values collective consumer/democratic processes, and may compromise on technical accuracy in the interest of "democratic" interests.

Methodology

The types of methodologies available to evaluators are numerous and will be influenced by the overall approach taken. Each has an extensive domain of knowledge and body of research. Examples include but are not limited to:

- Action research,
- Biography,
- Case study (individual, comparative),
- Economic evaluation (cost-minimization, cost-benefit, cost-effectiveness, cost-utility),
- Ethnography,
- Experimental,
- Quasi-experimental,
- Phenomenology, and
- Program theory.

Methods

Methods may be classified as quantitative, qualitative or mixed, the latter involving a combination of quantitative and qualitative methods.

Quantitative analyses may involve straightforward frequency counts and percentages or may include more involved statistical methods of analysis.

Qualitative methods include but are not limited to:

- Descriptive analysis based on observational study (e.g., narrative comparison of attributes),
- Interviews,
- Focus groups,
- Delphi methods,
- Discursive (text) analysis,
- Document review,
- Critical appraisal, involving in-depth analysis or perspective of an expert, and
- Storytelling (narrative, photovoice).

Product considerations

Report requirements and assessment of quality

Report requirements are generally established at the outset and may be stipulated in the contract between the evaluator and the sponsor. The types of requirements that the sponsor may want to make explicit include the audience, organization or presentation of findings, length, style of a final report. As well, depending on the evaluation timeframe, the sponsor may wish to request an interim report in which results from the earlier phases of the evaluation are reported.

In an effort to facilitate optimal practice in the field of evaluation, several evaluation associations offer standards or guidelines to assist program evaluators in conducting and reporting their evaluations. In the United States, the Joint Committee on Standards for Educational Evaluation has produced a series of standards that outline accepted practice for evaluators.[21] These standards have been adopted by the American Evaluation Association, one of the organizations represented on the Joint Committee. The German Evaluation Society (DeGEval) has posted similar evaluation standards on its website. Both organizations suggest that all evaluations should possess four fundamental attributes and have organized their standards according to these attributes: utility, feasibility, propriety and accuracy. These may be used by evaluation sponsors to guide their expectations of the evaluation and the assessment of the quality of the evaluation report.

The AEA and DeGEval summaries of standards are reproduced with permission in Appendix D.

In addition to the standards, Western Michigan University's Evaluation Center has published a series of checklists on their website (www.wmich.edu/evalctr/checklists). Checklists are available to guide evaluation processes, such as developing an evaluation design and items that should be included in evaluation reports.

Evaluation outcomes

Sponsors and evaluators may wish to track and report the impact or disposition of the evaluation on the agency under study. This step fosters an ability to determine the effectiveness and value of the evaluation. The categories of outcomes suggested through our review process were:

- Internal improvements – those applied with a view to ongoing development and enhancement to the program. This is consistent with the concept and intentions of a formative evaluation.
- Decision regarding the future of the agency – indication that decisions with respect to continuation or discontinuation of the agency resulted specifically from the findings of the evaluation.

- Minimal or no effect – there is no indication that any changes were made to the program or agency, based on the results of the evaluation.

REVIEW FINDINGS

In this section, the results of the review of HTA evaluations are presented, organized by the elements of the conceptual framework.

General

Cost

Of the nine evaluations for which cost information (in US dollars) was made available, four were estimated to cost between $50,001–100,000; three were estimated to cost between $20,001– 50,000; and two were reported to cost less than $20,000. The budget of the agencies are unknown, therefore the cost as a percent of agency budget was not calculated.

Availability

Eight of 16 evaluation reports were published; of these, four are available on a website. Three were reported to be available upon request. Five of the 16 were confidential and available only to selected readers.

Part of a larger review

Six evaluations were conducted as components of wider reviews, including a review of the role of the Ministry of Health in supporting HTA groups, site visits to a selection of innovative general practice commissioners, the evaluation of collaborative agreements between health service authorities, and complementary internal and external HTA reviews.

Findings

The findings and conclusions in the evaluations were by and large specific to the HTA agency studied, and few results are considered transferable to other agencies.

Generally, evaluations were favourable to the HTA agencies, although most contained numerous recommendations, mainly for quality improvement.

Of potentially most use to other jurisdictions are two studies that attempted to directly evaluate impact – one on clinical practice and the second on government policy and health system costs.[6,11]

Sponsor

Ten of the evaluations were commissioned by government; the remaining six by the HTA agencies, generally in collaboration with government as funder or because of a contractual commitment to government.

Design considerations

Intent

Table 3 provides a summary of the reviewers' interpretation of the underlying, generally unstated, intent of the evaluations according to the characteristics listed in Table 1 above. Table 4 provides a detailed summary of the intent, stated purpose and target audience of each evaluation.

Many of the evaluations appeared to have more than one underlying driver. In these instances, both primary and secondary intent was coded.

Most of the evaluations were undertaken for accountability purposes. In three of these evaluations, sustainability of the agency appeared to be in question, and was possibly a concern in an additional two evaluations.

The next most prevalent intent was quality improvement, followed by problem solving and curiosity. In five evaluations where accountability was the apparent intent of the evaluation, the reviewers judged the actual "flavour" of the evaluation to more closely reflect an interest in quality improvement. This determination was based on the observation that no question of agency sustainability was mentioned and the evaluations, while being retrospective, were forward-looking with a clear focus on strengthening the organization for the future.

Table 3: Number of evaluations demonstrating primary and secondary intent

Intent	Objective	Primary Intent	Secondary Intent	Total
Accountability	Account for investment or determine if contractual obligations met	6	7	13
Improvement	Strengthen agency in future	7	2	9
Problem solving	To address specific issues	2	2	4
Curiosity	Generate knowledge	1	1	2

Table 4: Intent, purpose and target audience of evaluations

Sustainability an issue?*	Intent *		Themes in stated purpose	Target audience
	Primary	Secondary		
Y	Accountability		Pilot project Longer term role	Gov't & users
Y	Accountability		Value for money Progress in introducing evidence-based health care	Gov't
Y	Accountability	Improvement	Assist gov't in deliberations of agency's future	Gov't
Possibly	Accountability		Contractual requirement/renewal	Gov't
Possibly	Accountability	Curiosity	Value for money Understand changes in purchasing patterns or clinical practice	Gov't
N	Accountability	Improvement	Contractual requirement/renewal	Gov't & agency
N	Improvement	Accountability**	Contractual requirement/renewal	Gov't & agency
N	Improvement	Accountability**	Future development of agency Decision to evaluate established at outset in strategy document	Gov't & users
N	Improvement	Accountability**	Contractual requirement Determination of forward strategy	Gov't
N	Improvement	Accountability**	Support optimal future development Value for money	Gov't & agency
N	Improvement	Accountability**	Contract requirement/renewal	Gov't
N	Problem solving	Accountability	Issues: Board representation, funding formula Value to government and health system	Gov't
N	Problem solving	Accountability	Issues: Prioritization, variation in access to technology, independence Progress towards goal achievement	Gov't
N	Improvement	Problem solving	Assist agency to more effectively carry out mandate Issues: Location/governance; roles, responsibility & relationships	Agency
N	Improvement	Problem solving	Review of methods and processes Focus on controversial issues	Agency
N	Curiosity		Understand impact	Gov't & agency

* As judged by reviewer

** In these instances, accountability was the explicit or apparent intent based on the stated purpose. However, the reviewer's judged these evaluations as more clearly demonstrating other underlying intents.

Government was the apparent sole or primary target audience in eight evaluations. Four evaluations targeted both government and agencies, and two targeted primarily or solely the HTA agency. Two included both government and potential users (e.g., clinicians) as target audiences.

The following themes were stated as driving forces or purposes of the 16 evaluations:

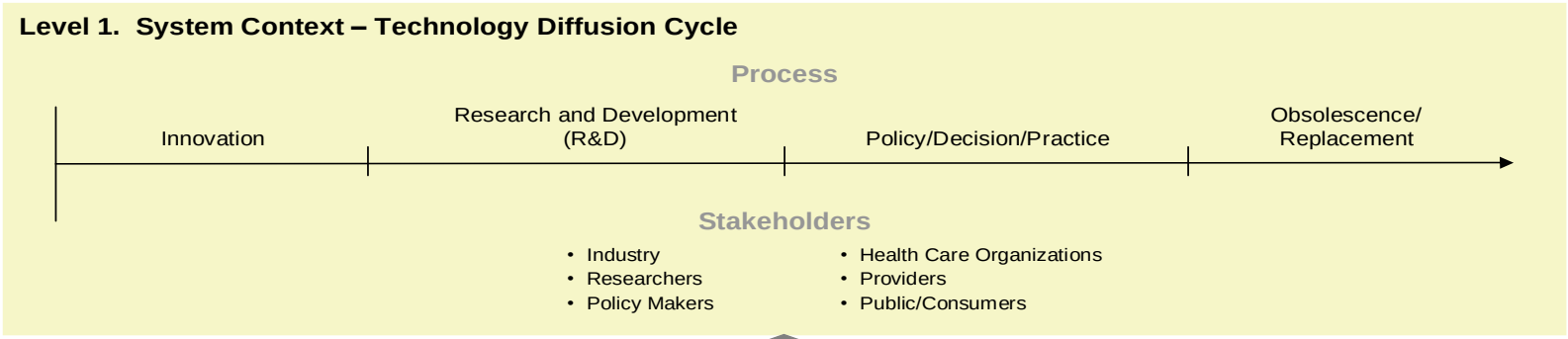
- Five evaluations were prompted by a contractual requirement or renewal;
- Five mentioned interest in the future direction or development of the agency;

- Four mentioned the need to determine value for money;
- Four noted specific issues to be addressed; and
- Two included an interest in understanding impact as a purpose for the evaluation.

Dimensions evaluated

In Figure 4, the number of evaluations that targeted specific dimensions of the HTA agency model is highlighted in red.

Figure 4: Dimensions evaluated



[7]

(Adapted from Menon D, Stafinski T [20])

Level 2. HTA Agency

Inputs/Structures [12]	Processes [9]	Outputs [9]	Impact of Products [10]	Ultimate HTA Outcomes [2]
▪ Mandate [2] - Target audience ▪ Principles/values [2] ▪ Governance [6] ▪ Contractual relationship [5] ▪ Collaborative relationships [5] ▪ Financial resources [6] ▪ Staff/human resources [4] ▪ Committee/board structures/roles [3] ▪ Organizational structure [4] ▪ Data/information systems [1] ▪ Target audience [1] ▪ Sources of requests [1]	Management processes: [2] ▪ HR management ▪ Financial management ▪ Project management ▪ Strategy/planning [1] ▪ Evaluation &/or research [1] ▪ Communications [1] HTA processes: ▪ HTA selection & prioritization [7] ▪ HTA question formulation ▪ Commissioning & monitoring [1] ▪ Data collection & analysis [1] ▪ Decisions/recommendations [2] ▪ Report preparation & review [2] ▪ Dissemination/research transfer [4] ▪ Appeals [2]	Horizon scans HTA products (full): ▪ Description [4] ▪ Quality [6] ▪ Cost [2] ▪ Timelines [1] ▪ Relevance [1] Appraisals [2] Recommendations [1] Other ▪ Research transfer or capacity building events/products	Acceptance of agency or products ▪ Awareness [4] ▪ Attitude [9] ▪ Satisfaction [3] Utilization of HTA products: ▪ Symbolic ▪ Conceptual – change in awareness, knowledge, attitude about the technology ▪ Instrumental – change in policy or practice [7] Impact on technology: ▪ Innovation or adaptation ▪ Research & development ▪ Obsolescence/replacement	Impact on health status [1] Impact on health system ▪ General [1] ▪ Economic/System costs [1] ▪ Equity ▪ Sustainability

(Adapted from Hailey[3])

Legend: [x] = number of evaluations

Context

Seven evaluations considered contextual issues, such as the placement or role of the agency within a wider context, role of the agency in national and international collaboration, role in implementation of results with respect to policy decisions and clinical guidelines, relationship of the agency with the pharmaceutical industry, stakeholder expectations, and government needs and extent that agency could meet this need.

Inputs/Structures

Twelve evaluations addressed aspects of inputs/structures. Financial resources or budget, governance, contractual and collaborative relationships, staffing, organizational structure and committee/board roles received the most attention in this category.

The evaluations addressing staff or human resources looked at the skills or competency, utilization of staff resources or who should undertake the analyses, and level of support resources. One addressed management skills.

The evaluations that considered committee structures or functions were interested either in the support provided by management committee, functions of the appraisal committee or role of the advisory, scientific and editorial boards.

Two evaluations examined the organization's principles or values. Transparency, inclusiveness, credibility, independence, evidence-based, consultation and responsiveness to change were principles mentioned.

Processes

Nine evaluations included a focus on processes. The HTA processes receiving the most attention in the evaluations were HTA selection and prioritization, assessment framework/process, and dissemination. Those looking at the assessment framework or process considered the development and use of a sound framework, consistent use of the framework, and quality of analysis. Those evaluating dissemination processes considered the role of the agency in dissemination, information/promotion strategies and international publication of results.

Management processes received less attention, but included consideration of the formation of commissioning and communication, setting up expert groups and strategy and research.

Outputs

Nine evaluations considered the agency's HTA products. This included description of the products and/or assessment of their quality, format, consistency, relevance, timeliness and cost. Two output evaluations employed a case study methodology whereby the quality of selected HTA reports was explored in depth.

Impact

Four evaluations addressed awareness of the agency or its products; nine addressed attitude and three looked at satisfaction. Attitude included perceived value, benefit, relevance or utility of products (considered in two evaluations) or of the agency (five evaluations), or general reputation of the agency or products (one evaluation).

Seven evaluations considered instrumental utilization of products. With two exceptions, impact was measured indirectly by asking stakeholders their perceptions of utilization or implementation of recommendations. Two attempted a direct evaluation of impact on or policy or practice change, or more general utilization.

Ultimate Outcomes

Two studies considered ultimate impact. In one case, the review addressed this question indirectly by asking survey respondents their perception of impact on health status or the health system. The other study undertook a direct assessment of impact on health care costs.

Evaluation approach and evaluator type

Four evaluation approaches were evident in the evaluations (Table 5). Ten took a program evaluation approach, although the scope, scale and rigour of the evaluations varied considerably. Generally, they involved the formation of evaluation questions, either pre-determined by the sponsor or determined collaboratively between the sponsor, agency and evaluator.

Three evaluations could be described as applying predominantly the criticism and connoisseurship approach described as by Stufflebeam.[1] In one of these cases, the evaluation team appeared to have been given considerable freedom to design the evaluation, including the formation of the evaluation questions to be addressed. In two of these three evaluations, the findings of stakeholder interviews and other data collection methods were not distinguished from the authors' findings and conclusions.

Two evaluations involved a research approach. The final, and perhaps most interesting approach, was an evaluation undertaken by parliamentary committee that applied a judicial approach whereby the committee heard testimony from a range of stakeholders on the merits and issues of the program.

Table 5: Evaluation approach and evaluator type

Approach	Evaluator Type	Number
Program evaluation	Program evaluators	4
	Government staff	3
	Content expert	2
	Researchers and content experts	1
Constructive critic	Researchers and content experts	3
Research	Researcher	1
	Government staff	1
Judicial	Parliamentary committee	1

The large majority of evaluators or evaluation teams (13/16) were assigned or pre-determined by the HTA agency, government sponsor or both. Some evaluations noted that assignment was based on expertise, skills and previous experience of the evaluators. Of the 13 assigned evaluators, four were assigned to internal government staff members or teams. Three of 16 evaluators were selected via a competitive selection process.

Methodology and Methods

Table 6 presents the types of data collection methods used in the 16 evaluations. The primary methods used were qualitative, involving interviews/focus groups or written survey.

Table 6: Methodologies/methods

Data Collection/Analysis*	Number
Interview/focus group	14
Document review	13
Written survey	6
Observation	4
Case study	3
Other: • Oral evidence • Self-evaluation • Utilization data • Web-site data • Environmental scan • Site visits • Comparison of staff and evaluator's SWOT analysis • Cost data	8

* Terms used are those of the report authors when describing their methodologies and methods.

Four of the 16 evaluations included either a formal or informal literature review. Several used the literature search as a means for identifying data or trends which they used for benchmarking the HTA agency with others agencies internationally.

Of the eight evaluations that described their data analysis, six undertook mixed quantitative-qualitative analyses; the remaining two undertook qualitative analyses only.

Product considerations

None of the evaluations reviewed described the sponsor's report requirements or assessment of quality nor was this information solicited in the validation process as this aspect was not a focus of the study.

Evaluation outcomes

Table 7 provides a brief overview of the known effects of the evaluation on the HTA agencies or programs, based on the content of the evaluation report or subsequent correspondence with an agency representative. The effect of the evaluation on the agency is known for 14 of the 16 agencies reviewed.

Table 7: Evaluation outcomes

Effect of Evaluation on Agency/Program	Number
Internal improvements	6
Decision regarding future of agency • Continued funding where sustainability was issue • Discontinuation • Broadened scope • Confirmed location	5
Minimal effect	3
Unknown	2

Six of the 16 evaluations reportedly resulted in internal improvements to the HTA agency, with a view to enhancing its structure and/or processes for the future. In the five cases where the evaluation was undertaken for the purpose of contract renewal or decision following pilot testing, the agencies' funding was renewed or continued.

In five cases, the evaluation appears to have contributed to a decision about the future of the agency. In two of the three cases where sustainability was an apparent issue, a decision was made to continue the service. In one instance, the agency was discontinued. In another, the evaluation resulted in a broadened scope with increased methodological research and more contact with the primary care sector. In the fifth case, a decision was made to continue the location of the agency under the existing

governing body contrary to a previous recommendation. While this did not represent a change to the organization, a decision to change locations may have impacted the sustainability of the agency in the long run.

In three instances, the evaluation appears to have had minimal direct impact on the agency.

DISCUSSION

The 16 evaluations reviewed represented a wide variety of approaches to the evaluation of HTA agencies or programs and offered the opportunity to learn from these differences while testing a conceptual framework for evaluation. Several limitations of the review are noted for the reader's consideration.

As reviewers, we had access to the written reports supplemented by written feedback received through the extraction table validation process. However, we know that the complete story of each evaluation includes undocumented information and nuances to which we were not privy. Ideally, we would have gained valuable insight through individual interviews with the evaluation sponsors and senior executives of the HTA agencies, provided they would have been willing to share their perceptions and more detailed knowledge of the evaluation. In the absence of this, we could rely only on the written report, and in some cases, use our judgment to read between the lines; for example, to ascertain the underlying intent of the evaluation.

In selecting reports for review, we included those entitled "reviews" as well as those entitled "evaluations" as we wished to describe the full range of evaluative activity undertaken of HTA agencies in the last decade. In the field of program evaluation, "evaluation" may be defined as "the systematic collection and analysis of information on the performance of a policy, program or initiative to make judgments about relevance, progress or success and cost-effectiveness and/or to inform future programming decisions about design and implementation".[22] On the other hand, "reviews are often conducted in response to a pressing or immediate need of management. As such, the emphasis is usually on quick generation of sufficient information to inform decision-making or reassure senior management of the dimensions of a problem or situation. The methodology used to gather information is usually secondary to developing an adequate answer in a timely fashion."[23] Upon review of the reports, it became clear that several of the initiatives represented reviews rather than evaluations of the HTA agency. This was especially true of the reports whose underlying intent was judged to be "problem-solving/issue resolution". Because reviews should not be held to the same standards or guidelines as evaluations, and do not necessarily fit well into an evaluation framework, we recommend that any future similar review exclude such reviews in the analysis.

In the reporting of our findings, we honoured our commitment to several agencies to ensure anonymity in the findings. This limited our ability to provide details of the evaluations and their findings.

Six of the 16 evaluations were of Canadian agencies and, as a result, Canada is over-represented in the sample. This may be attributed to the fact that we were most familiar with evaluations undertaken in Canada. Alternatively, it may reflect a longer time horizon for the Canadian agencies or greater interest in evaluation.

Despite the limitations, several patterns emerged from this review. First, it appeared that the underlying intent of the evaluations determined other decisions, such as the evaluator type, evaluation perspective or orientation and aspects evaluated. Of note is that this intent was generally not explicit in the evaluation report, and in some cases, the statement of purpose (reflecting accountability) was not borne out by the results and other characteristics of the evaluation, which clearly reflected a quality improvement intent. We noted several instances where the evaluation sponsors appeared to use the contractual requirement for evaluation as an opportunity to provide forward looking recommendations for enhancement rather than to seriously critique past performance with a view towards a summative decision about continuation of funding.

Second, it was interesting (although perhaps not unexpected) to observe a preference for assessment of structures, processes and outputs over evaluation of impact or outcome. This may reflect the fact that as a whole the agencies represented in our sample were relatively young at the time of their evaluation and formative reviews may have been more appropriate than other types of evaluation. Alternatively, it may reflect the fact that formative evaluations are easier to conduct within a shorter time frame and less expensive than rigorous evaluations of impact. Eight years have elapsed since Jacob and McGregor observed that "failure to make any attempt to assess the impact of HTAs is clearly inconsistent with the concept that decisions should be evidence-based" (p. 68).[6] It will be interesting to observe whether HTA agencies move towards greater emphasis on impact assessment as they mature.

Third, we note that few evaluations appeared to be informed by the literature or by other evaluations, likely related to the fact that the evaluations are generally not published for confidentiality reasons. Only two of the 16 evaluations identified in this study were reported in the peer reviewed literature. An interesting question arises as to why this might be the case. Are there barriers or lack of interest on the part of publications to accept and evaluators to submit such evaluations? Were the evaluations not of sufficient rigor to merit passing the peer review process, or do agencies generally perceive that evaluation is an internal organizational issue with limited applicability or interest for others? Does this reflect the fact that the majority of evaluations were sponsored by government rather than by HTA agencies? Perhaps there is a perception that the material reported therein should not be made available for public scrutiny. Should evaluation of agencies be held to the same transparency principle as HTA products themselves? We suggest the need for more activity on the part of HTA agencies and their funding partners to publish the results of their evaluations, in the interest of transparency.

Fourth, despite the limitation of not knowing the nuances of each evaluation, we observe the importance of considering the socio-political context of the HTA agency when undertaking an evaluation. While not included in this review due to our timeline inclusion criterion, we note that the abolition of both the U.S. Office of Technology Assessment and British Columbia Office for Health Technology Assessment occurred

following evaluation activity in each agency. In addition, one of the agencies included in this review was discontinued following review. However, upon review of the circumstances, it is apparent that in each case, the evaluation had limited impact on the closure decision. In evaluating health technology agencies, evaluators must be cognizant of the socio-political forces at play. Speaking to evaluators, Chelimsky (1997) advocates that "it is time to recognize . . . that our ability to serve policy depends as much on what we understand about how politics works as it does on the quality and appropriateness of our methods" (p. 55).[24]

REFLECTIONS

This exploratory study is a first step towards better understanding the evaluations undertaken of HTA agencies. Our concluding comments, organized under the two study objectives established at the outset are offered to assist HTA agencies contemplating evaluation activity.

Objective 1: To propose a generic evaluation framework for HTA agencies to strengthen their evaluation capacity.

A conceptual framework was developed, tested and refined based on the findings of this review. It is hoped the framework offers a systematic way of designing and undertaking future evaluations of HTA agencies. In addition, HTA agencies are encouraged to access and make use of readily available program evaluation standards, guidelines and checklists that could prove useful for conducting their own internal evaluations, or for monitoring and assessing the quality of the evaluations they commission.

Given the needs and underlying intentions of various audiences and sponsors, different stages of evolution of the agency, and different perspectives in the various evaluator types, it appears that a one-size-fits-all evaluation design is not particularly useful. The conceptual framework, therefore, proposes a series of options for consideration, according to the various design, review and follow-up considerations involved in conducting an evaluation. Further research aimed at understanding the relationships between various aspects of the conceptual framework would serve to provide more specific guidance to sponsors and evaluators.

Objective 2: To conduct a review of HTA agencies to understand what aspects of HTA agencies have been evaluated, approaches/methods used, outcomes of the evaluations and to understand what was learned through these evaluations that could guide HTA agencies to best serve their mandate.

The review revealed a variety of approaches used by sponsors and evaluators, based on the underlying intent of each evaluation. The aspects evaluated include the context of the agency, its structures, processes, products and impacts with the majority of evaluators focusing on the structures, processes and products. Most evaluations relied on qualitative methods involving interviews and focus groups where perceptions of stakeholders were solicited. Only two used a more rigorous approach to assessing impact.

The findings and conclusions in the evaluations were, by and large, specific to the HTA agency studied, and few results are considered transferable to other agencies. Generally, evaluations were favourable to the HTA agencies, although most contained numerous recommendations, mainly for quality improvement.

A number of possible guiding principles for the evaluation of HTA agencies are offered based on the findings of this review.

1. We advocate for a more purposeful evaluation design, whereby sponsors and evaluators consider the alternative approaches, based on their fundamental intent for the evaluation. The sponsor and evaluator should be explicit about the main intent and purpose of the evaluation.
2. We suggest that future evaluations make use of and build on previous knowledge and experience gained through other evaluation. Sponsors may wish to conduct a literature search to contextualize and build on the work of other agencies and evaluators. The literature search should include the HTA and evaluation bodies of literature.
3. We suggest that sponsors recognize that different types of evaluators bring different skills and approaches to the task. Different types of evaluators may be appropriate, depending on the main intent or purpose. Often the strongest approach will be one that includes evaluators of different backgrounds in an evaluation team.
4. We advocate that evaluators be deliberate and contextual about the methodological approaches employed.
5. We recommend that sponsors and HTA agencies/programs consider the question of whether the evaluation should be made public in an effort to be transparent and to help build the body of knowledge around the effectiveness of HTA agencies.
6. As a general rule, it is important that the merit or worth of an organization is not evaluated before it has been determined that the processes and inputs/structures have been implemented as planned, are adequate to achieve the intended outcomes and are functioning at an optimal level. That is, formative evaluation generally should precede impact or summative evaluation.

APPENDIX A: METHODOLOGY

The review was undertaken as a three-step process, as described below.

Literature search

A targeted literature search was undertaken with a view to developing a preliminary conceptual framework to guide the collection, compilation and analysis of the evaluation information. A second purpose of the literature review was to identify any evaluations for the review that may have been missed through the direct appeal to INAHTA members. The search targeted literature in both HTA and evaluation fields, published between 1990 and October 2004.

A research librarian conducted the literature search. Major electronic databases used include: The Cochrane Library, NHS Centre for Reviews and Dissemination (CRD) Databases: NHS EED, HTA, DARE, and PubMed. Hand-searching of *The Canadian Journal of Program Evaluation*, *The American Journal of Evaluation*, and *New Directions for Evaluation* was conducted to find relevant literature. In addition, relevant library collections, web sites of evidence-based resources, and HTA and HTA-related agency resources were searched. Internet search engines were also used to locate grey literature. Search terms listed by database and platform are presented in Appendix A.

Development of conceptual framework

We used an iterative process to achieve our dual purpose to develop a conceptual framework and undertake a review of evaluations. A preliminary framework was developed based on the findings of our literature search. The preliminary framework informed the organization and content of the data collection instrument for the review. Once the review of HTA evaluations was completed, the conceptual framework was revisited and refined based on the findings from the review. Then, the final framework was used to organize the reporting of the review results.

Review HTA evaluations

Inclusion criteria

The following inclusion criteria were applied in the selection of evaluations for review:

- The subject must be a public sector HTA agency defined as a member of INAHTA.
- The scope of the evaluation must be the entire HTA agency or a substantial core or central function of the agency.
- The evaluation must be conducted by an external, independent evaluator defined as an individual not employed by or affiliated with the agency at the time of the evaluation.

- The evaluation must have been completed between January 1995 and January 2005.
- The evaluation may have been undertaken as part of a broader review or evaluation, for example, of the organization of which the HTA agency is a part.
- Evaluations included those labeled as reviews.
- The evaluation may be either formative or summative.

Identification of potential evaluations

An initial request for English language evaluations, conducted between the years 1999 to 2004, was issued by electronic mail to INAHTA members in August 2004. This request was subsequently expanded to include all languages and the 10 years 1994 to 2004. A final tabulation of the agencies that had undergone an external review during the decade was shared with all INAHTA agencies with a request to identify any that were missed.

Twenty-one of 39 INAHTA members responded to the inquiry. Of these, 11 indicated that no evaluation had been conducted within the parameters of this study. An additional 10 stated that one or more evaluations had been conducted and shared these evaluations with the review team. Fourteen eligible evaluations were identified through this process. An additional two evaluations were identified through the literature search. They did not represent current INAHTA members but because they represented precursors to current INAHTA agencies, were included in the review. Thus, a total of 16 evaluations were deemed eligible and reviewed. These evaluations represent 12 HTA agencies located in eight countries.

One evaluation was translated into English for the purpose of the review. Three evaluations were shared with the investigators under the condition of confidentiality and anonymity.

Data collection, compilation and validation

A data extraction table was developed, incorporating aspects of the preliminary conceptual framework (Appendix B). With the exception of two items, the data extraction table was completed by one reviewer, and independently verified by a second member of the three person review team. The remaining two items – evaluation approach and aspects evaluated – involved judgment, and were independently assessed by two team members. Discrepancies in coding were resolved through discussion. Both reviewers also independently noted their observations about the evaluation.

Once completed, the data extraction tables were sent to the senior executive in each HTA agency for validation. The version validated did not include the reviewer's coding of "evaluation approach" nor their observations and notes. An additional section on cost of the evaluation was added, as no cost information was found in any of the evaluations reviewed. All 16 completed data extraction tables were validated and

returned by the HTA agency. Adjustments were made based on the comments and additions received. Generally, the adjustments involved minor clarifications (e.g., position titles, target audience) or elaborations on information not generally evident in the report (e.g., final outcome of the review, cost).

Final analysis and reporting

Once validated, the review results were synthesized by item across all evaluations. Upon completion of the synthesis, the reviewers revisited their preliminary conceptual framework in light of the review findings. They then re-analyzed and re-organized the findings of this review according to the final framework. The re-analysis did not change the validated information content.

Database	Platform	Date Searched	Search Terms †*
The Cochrane Library	Issue 3, 2004 Update Software	2004-10-15	(agenc* OR organisational OR organizational) NEAR (review* OR evaluation OR assess*) AND (HTA OR (health technology assessment))
CRD (UK): Health Technology Assessment Database NHS Economic Evaluation Database Database of Reviews of Effects	http://nhscrd.york.ac.uk	2004-10-18	independent evaluation OR external evaluation OR external review OR organization review OR organisation review
PubMed National Library of Medicine	http://www.pubmed.gov	2004-10-12	(agenc* OR organisational OR organizational) AND (review* OR evaluation OR assess*) AND HTA **OR** (program evaluation OR management audit) AND (government agencies OR health planning councils) AND (technology assessment OR HTA)
EMBASE	Ovid	2004-10-18	(independent evaluation or external evaluation or external review or organization$ review or organisation$ review).mp. AND (HTA or technology assess$).mp.

Database	Platform	Date Searched	Search Terms † *
Web of Science Science Citation Index Expanded (SCI-EXPANDED) Social Sciences Citation Index (SSCI) Arts & Humanities Citation Index (A&HCI)	ISI	2004-10-11	(independent evaluation* OR external evaluation* OR external review* OR organization* review* OR organisation* review*) AND (HTA OR technology assess*)
NEOS library catalogue	www.library.ualberta.ca	2004-10-11	program evaluat$ AND technolog$ assess$
Amicus – Library and Archives Canada	http://www.collectionscanada.ca/amicus/	2004-10-12	Title Keyword search: evaluat? technolog? assess?
US Library of Congress	http://catalog.loc.gov/	2004-10-12	Keyword search: "external evaluation" "health technolog? assess?"
Other web searching:	Google, Copernic Agent Basic, HTA and HTA-related agencies websites		
Handsearching of the following journals:	Canadian Journal of Program Evaluation American Journal of Evaluation New Directions for Evaluation		

† Limits**:** Searches were limited to publication dates: 1990-2004. The limits were applied in databases where such function is available.

* **Notes:** There are truncation symbols used in the searches, such as: $, *,?. A truncation character retrieves all possible suffix variations of the root word e.g. surg* retrieves surgery, surgical, surgeon, etc.

APPENDIX B: EXTRACTION TABLE TEMPLATE

Review of HTA Evaluations

Agency name:			
Country:			
Year Completed:			
Report Title			
Contact:			
URL:			
Commissioned by:			
Evaluator name:			
Position title:			
Educational qualification:			
Type:	Academic		
	Consultant		
	Not stated:		
	Other		*Describe:*
Selection process:	Competitive		
	Assigned		
	Not stated		
	Other		*Describe:*
Cost:			
	Not stated		
Report is:	Published		
	On web-site		
	Avail on request		
	Confidential		
	Unknown		
Part of larger review?	Yes		*Describe:*
	No		
What prompted review/evaluation?			
Was a literature review conducted?	Yes		*Describe:*
	No		
Overall purpose:			
Objectives:			

Evaluation questions:			
Target audience:	Stated Implied Cannot imply		*Describe:*
Reference to expectations, e.g., theory, model, targets, norms?	Yes No		*Describe:*
Aspects evaluated? (see attached)			
Evaluation approach: (see attached)	*List all*		*Describe:*
Data collection method:	Document review Interview Written survey Observation Case study Other		*Describe (by aspect evaluated, if appropriate):*
Analysis:	Quantitative Qualitative None described		*Describe:*
Stated limitations: (include report of conflict of interest)			
Findings/results (by component), including problems/issues, success factors, lessons learned			
Conclusions			
Recommendations/ Suggestions for improvement:			
Final outcome of review/evaluation:			

Reviewer notes, including quality issues, and limitations:	
Completed by:	Date:
Primary reviewer:	
Second reviewer:	

Aspects reviewed/evaluated (adapted from Hailey 2003[2])

- HTA Context
- HTA Inputs/Structures
  - Governance
  - Mandate
  - Collaborative relationships
  - Contractual relationship
  - Resources
  - Staff and structure
  - Access to data
- HTA Processes
  - Formulation of HTA question
  - HTA Evaluation
  - Provision of advice and dissemination
- HTA Products
- Impact
  - Direct
  - Indirect
- Ultimate impact – changes to health care or health outcomes

Evaluation approach (from Stufflebeam 2001[1])

Pseudo evaluations

1. Public Relations-Inspired Studies
2. Politically Controlled Studies

Questions- and Methods-Oriented Evaluation Approaches

3. Objectives-Based Studies
4. Accountability, Particularly Payment by Results Studies
5. Objective Testing Programs
6. Outcome Evaluation as Value-Added Assessment
7. Experimental Studies (controlled designs)
8. Performance Testing
9. Management Information Systems
10. Benefit-Cost Analysis Approach
11. Clarification Hearing
12. Case Study Evaluations
13. Criticism and Connoisseurship
14. Program Theory-Based Evaluation
15. Mixed-Method Studies

Improvement/Accountability-Oriented Evaluation Approaches

16. Decision/Accountability-Oriented Studies
17. Consumer-Oriented Studies
18. Accreditation/Certification Approach

Social Agenda/Advocacy Approaches

19. Client-Centered Studies (or Responsive Evaluation)
20. Constructivist Evaluation
21. Deliberative Democratic Evaluation
22. Utilization-Focused Evaluation

APPENDIX C: INAHTA IMPACT FORM

Framework for reporting on impact of HTA reports

To complete the form, please put the marker on the grey areas to write text and double click on the squares to tick them. Reference should be made to the accompanying instructions. The form should be sent to the INAHTA Secretariat via e-mail <<nordwall@sbu.se>>

A. Agency: [Write here]	**B. Name of technology:** [Write here]	B.1 [Add any needed qualification – eg particular application]
C. Date of this record: [month and year]* ***The date of the record should be not less than 6 months after the publication date of the HTA report***		**D. Date of HTA report:** [month and year]
E. Origin of request for the HTA: [Give the name or type of organization that made the request]		
F. Purpose of the HTA	F.1 [Tick] ☐1 Coverage decisions ☐2 Capital funding decisions ☐3 Formulary decisions ☐4 Referral for treatment ☐5 Program operation ☐6 Guideline formulation ☐7 Influence on routine practice ☐8 Indications for further research ☐9 Other [Write here]	F.2 [Explanation/qualification, if needed]
G. Conclusions reached by the HTA [1 or 2 sentences]		

H. Indications of impact	H.1. [Tick one or more] ☐1. HTA considered by decision – maker ☐2. HTA recommendations/ conclusions accepted ☐3. HTA demonstrated that technology met specific program requirements ☐4. HTA material incorporated into policy or administrative documents ☐5. HTA information used as reference material ☐6. HTA linked to changes in practice ☐7. No apparent impact ☐8. Other [Write here]	H.2 [1 or 2 sentences to give further information]
I. AGENCY'S opinion on level of impact of the HTA	I.1. [Tick 1] ☐1. No apparent influence ☐2. Some consideration of HTA by decision maker ☐3. Informed decisions ☐4. Major influence on decisions	I.2 [1 or 2 sentences indicating basis/ reasons for opinion]
	I.3 Indicate any unintended influence that the HTA had [Write here] Did the unintended influence lead to a change in HTA procedure? [Yes/ No]	
J. EXTERNAL opinion on level of impact of the HTA	**Source of opinion:** [Write here] [Tick 1] ☐1. No apparent influence [Write here] ☐2. Some consideration of HTA by decision maker [Write here] ☐3. Informed decisions [Write here] ☐4. Major influence on decisions [Write here]	

INAHTA – Framework for reporting on impact of HTA reports

Instructions for use

Framework section	Action	Comments
A. Agency	Enter the acronym or name of your agency in this box	
B. Name of technology	Enter the name of the technology that was considered by the HTA	
	In box B.1 add any further explanation of the technology, for example a particular application that was considered	Entry of such information is optional
C. Date of this record	Enter the date that this record (the impact framework) was completed	As indications of impact may take some time to become apparent, the date of the record should be at least 6 months after the publication date of the HTA report. 6 months is the minimum period. The timing of the record of impact after 6 months is a matter for the agency to determine.
D. Date of HTA report	Enter the date of publication of the HTA report	
E. Origin of the HTA request	Enter the name or the type of organization that made the request for the HTA. If the HTA report was not requested from outside your agency, please indicate this.	Organizations might be government – related (eg health ministries) or non – government (eg professional bodies).
F. Purpose of the HTA	In box F.1 are eight types of decision that might have been informed by the HTA. Please mark one or more of these, as appropriate.	If there was some other type of decision that was informed by the HTA please mark " #9 Other" and briefly mention what it was
	In Box F.2 add any explanation regarding the type of decision that seems appropriate	This is optional. One or two sentences would be sufficient.
G. Conclusions reached by the HTA	Briefly outline the conclusions reached by the HTA.	One or two sentences would be sufficient. If appropriate, these might include major recommendations that were made.

Framework section	Action	Comments
H. Indications of impact	In Box H.1 are seven possible indications of the impact the HTA might have had . Please mark one or more of these. If there was some other type of impact of the HTA please mark "#8 Other" and briefly mention what it was.	1. HTA considered by decision - maker. [The HTA was considered but further impact was not obvious/ apparent.] 2. Acceptance of HTA recommendations/ conclusions [clear acceptance of HTA findings possibly, but not necessarily, linked to action by the decision maker.] 3. HTA demonstrated that a technology met specific program requirements [in circumstances where the HTA and its findings are linked to a program, for example where minimum standards must be met before some type of approval is given.] 4. HTA material is incorporated into policy or administrative documents [Material in an HTA is cited in subsequent documentation.] 5. HTA information used as reference material. [The HTA is used by decision makers as an ongoing source of information] 6. HTA linked to changes in practice [The HTA may be one of a number of factors influencing such change] 7. No apparent impact
	In Box H.2 provide further information, as appropriate.	One or two sentences should be sufficient.
I. Agency's opinion on level of impact	In Box I.1. are four categories of influence of the HTA. Please mark one of these to indicate the opinion of your agency on the level of impact that was achieved.	
	In Box I.2 briefly indicate the basis for your agency's opinion	1 or 2 sentences should be sufficient Details might include reasons for the report having no apparent influence, or the way in which the agency's opinion had been formed (for example through a survey of stakeholders).
	If the HTA had an unintended influence, please note this in Box I.3 Also note if the unintended influence led to a change in HTA procedure at your agency	For example, the conclusions of the HTA might have been misunderstood by a decision maker and action taken that was contrary to the intent of the HTA. Reference could be made here to any significant media coverage that may have increased the impact of the HTA report.

Framework section	Action	Comments
J. External opinion on level of impact of the HTA	Please note the **source** of any external opinion on level of impact. Inclusion of this information is essential if this box is to be completed. Please mark one of the four possible categories of influence of the HTA. to indicate the opinion of other organizations on the level of impact that was achieved.	For example, feedback may have been obtained from the organization that requested the HTA. Organizations such as patients/consumer groups and professional bodies may also be sources of opinion on impact

APPENDIX D: EVALUATION STANDARDS

U.S. Joint Commission on Standards for Educational Evaluation

Reproduced with permission

THE PROGRAM EVALUATION STANDARDS

Summary of the Standards

Utility Standards

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

U1 Stakeholder Identification Persons involved in or affected by the evaluation should be identified, so that their needs can be addresse.

U2 Evaluator Credibility The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.

U3 Information Scope and Selection Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders

U4 Values Identification The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear

U5 Report Clarity Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.=

U6 Report Timeliness and Dissemination Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion

U7 Evaluation Impact Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased

Feasibility Standards

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

F1 Practical Procedures The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained

F2 Political Viability The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted

F3 Cost Effectiveness The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified

Propriety Standards

The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

P1 Service Orientation Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants

P2 Formal Agreements Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

P3 Rights of Human Subjects Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects

P4 Human Interactions Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed

P5 Complete and Fair Assessment The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed

P6 Disclosure of Findings The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results

P7 Conflict of Interest Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results

P8 Fiscal Responsibility The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate

Accuracy Standards

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

A1 Program Documentation The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified

A2 Context Analysis The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified

A3 Described Purposes and Procedures The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed

A4 Defensible Information Sources The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed

A5 Valid Information The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use

A6 Reliable Information The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use

A7 Systematic Information The information collected, processed, and reported in an evaluation should be systematically reviewed, and any errors found should be corrected

A8 Analysis of Quantitative Information Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered

A9 Analysis of Qualitative Information Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered

A10 Justified Conclusions The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them

A11 Impartial Reporting Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings

A12 Metaevaluation The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses

German Evaluation Society Standards

Reproduced with permission

Edited by: German Evaluation Society Deutsche Gesellschaft für Evaluation e.V. (DeGEval) / Wolfgang Beywl

Contact: DeGEval
Bücheler Weg 27
D-53347 Alfter

Tel: +49 228-96 49 99 09
Fax: +49 228-96 49 99 20
URL: http://www.degeval.de
Email: **standards@degeval.de**

The document Selected Comments to the Standards for Evaluation edited by German Evaluation Society / Wolfgang Beywl can be downloaded from www.degeval.de

Introduction

The following Evaluation Standards were ratified by the general assembly of the German Evaluation Society (Deutsche Gesellschaft für Evaluation- DeGEval) on October 4th, 2001. They are the result of a two-year discussion and preparation process which included a membership survey, an appointed Standards Committee, and a review process.

The twenty-five DeGEval-Standards are organized in four groups. This structure as well as many Standards, including titles and descriptive statements, were stimulated by the "Program Evaluation Standards" of the US-American "Joint Committee on Standards for Educational Evaluation" and adapted to the requirements of evaluation in Germany and Austria. The DeGEval-Standards were also inspired by the Swiss adaptation of the Joint Committee Standards which provides a generalization of these standards from educational to more diverse settings.

In its German original, this short version of the DeGEval-Standards is accompanied by a 30-page document which includes a clarification of the aims and the scope of the Standards, definitions of evaluation and other key concepts, an overview of different approaches to evaluation, comments on the application of the Standards, and a description of the development of the document itself as well as the review process.

By the end of 2004, the DeGEval-Standards will have been subjected to a second review process which will include research societies as well as professional associations. For more information, please see http://www.degeval.de or contact Wolfgang Beywl (wolfgang.beywl@univation.org), former Chair of the Standards Committee and board member of the German Evaluation Society.

Evaluations should feature four basic attributes:
Utility – Feasibility – Propriety – Accuracy.

Utility[*]

The Utility Standards are intended to ensure that an evaluation is guided by both the clarified purposes of the evaluation and the information needs of its intended users.

U1 Stakeholder Identification
Persons or groups involved in or affected by the evaluand should be identified, so that their interests can be clarified and taken into consideration when designing the evaluation.

U2 Clarification of the Purposes of the Evaluation
The purposes of the evaluation should be stated clearly, so that the stakeholders can provide relevant comments on these purposes, and so that the evaluation team knows exactly what it is expected to do.

U3 Evaluator Credibility and Competence
The persons conducting an evaluation should be trustworthy as well as methodologically and professionally competent, so that the evaluation findings achieve maximum credibility and acceptance.

U4 Information Scope and Selection
The scope and selection of the collected information should make it possible to answer relevant questions about the evaluand and, at the same time, consider the information needs of the client and other stakeholders.

U5 Transparency of Values
The perspectives and assumptions of the stakeholders that serve as a basis for the evaluation and the interpretation of the evaluation findings should be described in a way that clarifies their underlying values.

U6 Report Comprehensiveness and Clarity
Evaluation reports should provide all relevant information and be easily comprehensible.

U7 Evaluation Timeliness
The evaluation should be initiated and completed in a timely fashion, so that its findings can inform pending decision and improvement processes.

U8 Evaluation Utilization and Use
The evaluation should be planned, conducted, and reported in ways that encourage attentive follow-through by stakeholders and utilization of the evaluation findings.

Feasibility

The Feasibility Standards are intended to ensure that an evaluation is planned and conducted in a realistic, thoughtful, diplomatic, and cost-effective manner.

[*] Translation Wolfgang Beywl, Cologne, Germany and Sandy Taut, Los Angeles, USA – 02/11/2001

F1 Appropriate Procedures

Evaluation procedures, including information collection procedures, should be chosen so that the burden placed on the evaluand or the stakeholders is appropriate in comparison to the expected benefits of the evaluation.

F2 Diplomatic Conduct

The evaluation should be planned and conducted so that it achieves maximal acceptance by the different stakeholders with regard to evaluation process and findings.

F3 Evaluation Efficiency

The relation between cost and benefit of the evaluation should be appropriate.

Propriety

The Propriety Standards are intended to ensure that in the course of the evaluation all stakeholders are treated with respect and fairness.

P1 Formal Agreement

Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or to renegotiate it.

P2 Protection of Individual Rights

The evaluation should be designed and conducted in a way that protects the welfare, dignity, and rights of all stakeholders.

P3 Complete and Fair Investigation

The evaluation should undertake a complete and fair examination and description of strengths and weaknesses of the evaluand, so that strengths can be built upon and problem areas addressed.

P4 Unbiased Conduct and Reporting

The evaluation should take into account the different views of the stakeholders concerning the evaluand and the evaluation findings. Similar to the entire evaluation process, the evaluation report should evidence the impartial position of the evaluation team. Value judgments should be made as unemotionally as possible.

P5 Disclosure of Findings

To the extent possible, all stakeholders should have access to the evaluation findings.

Accuracy

The Accuracy Standards are intended to ensure that an evaluation produces and discloses valid and useful information and findings pertaining to the evaluation questions.

A1 Description of the Evaluand

The evaluand should be described and documented clearly and accurately, so that it can be unequivocally identified.

A2 Context Analysis

The context of the evaluand should be examined and analyzed in enough detail.

A3 Described Purposes and Procedures

Object, purposes, questions, and procedures of an evaluation, including the applied methods, should be accurately documented and described, so that they can be identified and assessed.

A4 Disclosure of Information Sources

The information sources used in the course of the evaluation should be documented in appropriate detail, so that the reliability and adequacy of the information can be assessed.

A5 Valid and Reliable Information

The data collection procedures should be chosen or developed and then applied in a way that ensures the reliability and validity of the data with regard to answering the evaluation questions.

A6 Systematic Data Review

The data collected, analyzed, and presented in the course of the evaluation should be systematically examined for possible errors.

A7 Analysis of Qualitative and Quantitative Information

Qualitative and quantitative information should be analyzed in an appropriate, systematic way, so that the evaluation questions can be effectively answered.

A8 Justified Conclusions

The conclusions reached in the evaluation should be explicitly justified, so that the audiences can assess them.

A9 Metaevaluation

The evaluation should be documented and archived appropriately, so that a metaevaluation can be undertaken.

REFERENCES

1. Stufflebeam C. Evaluation models. *New Directions for Evaluations* 2001;89:7-98.
2. Hailey D. *Elements of effectiveness for health technology assessment programs*. Alberta Heritage Foundation for Medical Research, editor. Edmonton, AB: Alberta Heritage Foundation for Medical Research; 2003. Initiative #9.
3. Rush B, Ogborne A. Program logic models: expanding their role and structure for program planning and evaluation. *Canadian Journal of Program Evaluation* 1991;6(2):95-106.
4. Hailey DM, Cowley DE, Dankiw W. The impact of health technology assessment. *Community Health Studies* 1990;14(3):223-34.
5. Drummond M. Evaluation of health technology economic issues for health policy and policy issues for economic appraisal. *Social Science & Medicine* 1994;38:1593-600.
6. Jacob R, McGregor M. Assessing the impact of health technology assessment. *International Journal of Technology Assessment in Health Care* 1997;13(1):68-80.
7. Lehoux P, Battista RN, Lance JM. Monitoring health technology assessment agencies. *Canadian Journal of Program Evaluation* 2000;15(2):1-33.
8. Gerhardus A. *Assessing the impact of HTA reports: a toolkit approach.* Presentation at Health Technology Assessment International, 2nd Annual Meeting June 20-22, 2005, Rome, Italy. 2005.
9. Battista RN, Feeny DH, Hodge MJ. Evaluation of the Canadian Coordinating Office for Health Technology Assessment. *International Journal of Technology Assessment in Health Care* 1995;11(1):102-16.
10. Jacob R, Battista RN. Assessing technology assessment. Early results of the Quebec experience. *International Journal of Technology Assessment in Health Care* 1993;9:564-72.
11. Dixon S, Coleman P, Nicholl J, Brennan A, Touch S. Evaluation of the impact of a technology appraisal process in England: the South and West Development and Evaluation Committee. *Journal of Health Services Research and Policy* 2003;8(1):18-24.
12. Lehoux P, Tailliez S, Denis JL, Hivon M. Redefining health technology assessment in Canada: diversification of products and contextualization of findings. *International Journal of Technology Assessment in Health Care* 2004;20(3):325-36.
13. Menon D, Topfer LA. Health technology assessment in Canada. A decade in review. *International Journal of Technology Assessment in Health Care* 2000;16(3):896902.

14. Cumper GE. *The Evaluation of National Health Systems*. New York, NY: Oxford University Press; 1991.

15. Kaplan RS, Norton DP. *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press; 1996.

16. Menon D, Stafinski T. *A Canadian health technology assessment strategy: a discussion paper*. Edmonton, AB: Prepared for the Federal/Provincial/Territorial Advisory Committee on Information & Emerging Technologies; 2003. Unpublished.

17. Treasury Board of Canada Secretariat. Lexicon for results-based management and accountability. Available from: *http:/www.tbs-sct.gc.ca/eval/pubs/RMAF-CGRR/rmafcgrr05_e_asp* 2002 (accessed 2005 Sep 2).

18. Lavis JN. A political science perspective on evidence-based decision-making. In: *Using Knowledge in Health Care: Multidisciplinary Perspectives*. Toronto, ON: University of Toronto Press; 2004: p.70-85.

19. Kunkle GC. New challenge or past revisited? The Office of Technology Assessment in historical context. *Technology in Society* 1995;17(2):175-96.

20. Morgan MG. Death by Congressional ignorance: how the Congressional Office of Technology Assessment - small and excellent - was killed in the frenzy of government downsizing. Available from: *http://www.wws.princeton.edu/~ota/ns20/ota95_f.html* 1995 (accessed 2005 Sep 2).

21. Joint Committee on Standards for Educational Evaluation. *The Program Evaluation Standards*. Thousand Oaks, CA: Sage Publications; 1994.

22. Treasury Board of Canada Secretariat. Guide for the development of results-based management and accountability frameworks. Annex A. Available from: *http://www.tbs-sct.gc.ca/eval/pubs/RMAD-CGRR/rmafcgrr05_e asp* 2002 (accessed 2005 Sep 2).

23. Treasury Board of Canada Secretariat. Evaluation guidebook for small agencies. Available from: *http://www.tbs-sct.gc.ca/eval/dev/sma-pet/guidelines/guidebook02_e asp* 2005 (accessed 2005 Sep 6).

24. Chelimsky E. The political environment of evaluation and what it means for the development of the field. In: *Evaluation for the 21st Century*. Thousand Oaks, CA: Sage Publishers; 1997.