Blueprints for informed policy decisions: A review of laws and policies requiring routine evaluation

Report from Kunnskapssenteret (Norwegian Knowledge Centre for the Health Services)
No 16–2012
Review

kunnskapssenteret
Norwegian Knowledge Centre for the Health Services

Background: Countries are committed to improving the health and welfare of their populations. Yet, we found only five examples of laws and policies requiring routine evaluation of public programmes. This suggests that the majority of countries and international organisations may not be fulfilling their political and ethical obligations to use well de-signed evaluations of policies and programmes routinely to inform decisions about how best to use available resources to achieve societal goals. It is possible, however, that existing laws and policies do not adequately reflect the degree to which such actions are already being undertaken. **Lessons:** A number of important lessons can be drawn from the experiences discussed in this review, including: • The enactment of laws and policies supporting routine use of evaluation can: Capitalise on broad reforms (e.g. those focusing on accountability and transparency), Build on existing laws and policies and existing evaluation experience, Be championed by a wide range of people, including auditors general, budget bureaus, multilateral organisations and donors, legislative branches of government, and heads of state

(continued)

Norwegian Knowledge Centre for the Health Services (Kunnskapssenteret)
PO Box 7004, St. Olavs plass
N-0130 Oslo
(+47) 23 25 50 00
www.kunnskapssenteret.no
Report: ISBN 978-82-8121-505-4 ISSN 1890-1298

no 16–2012

kunnskapssenteret

(continued from page one) • Linking evaluation and monitoring objectives to other government initiatives and institutions can create synergies in budgetary processes and accountability and transparency, in ways that create an integrated rather than overly-regulated system • Implementation of a monitoring and evaluation system goes hand in hand with administrative reforms. Such changes enable those responsible for monitoring and evaluation to respond to the information needs of decision-makers and to link monitoring and evaluation to decision-making • It is important to focus clearly on assessing the performance of programmes in order to identify the core issues that need to be addressed in evaluations (e.g. effectiveness, efficiency and equity), and to make clear what type of evidence is wanted • It is advisable to have an entity with a clear mission to carry out independent, unbiased evaluations to a high standard. The entity should be insulated from the influence of political organisations or interest groups.

Title Blueprints for informed policy decisions: A review of laws and policies requiring routine evaluation

Norwegian title Internasjonale erfaringer med lovpålagte krav til rutinemessig evaluering

Institution Norwegian Knowledge Centre for the Health Services
(Nasjonalt kunnskapssenter for helsetjenesten)
Magne Nylenna, *Director*

Authors Kebede, Mekdes, *PhD candidate, University of Oslo*
Oxman, Andrew David, *Senior researcher, Norwegian Knowledge Centre for the Health Services*
Glenton, Claire, *Senior researcher, Norwegian Knowledge Centre for the Health Services*
McKee, Martin, *Professor, London School of Hygiene & Tropical Medicine*
Mills, Edward J, Associate Professor, *Faculty of Health Sciences, University of Ottawa*
Pang, Tikki, Visiting Professor, *Lee Kuan Yew School of Public Policy, National University of Singapore*
Whitworth, Judith, *Professor, John Curtin School of Medical Research, Australian National University*

ISBN 978-82-8121-505-4

ISSN 1890-1298

Report No. 16 – 2012

Project number 624

Type of report Review (Oversikt)

No. of pages 62 (121 including appendices)

Client WHO Advisory Committee on Health Research and Norad

Subject heading (MeSH) *Evaluation Studies as Topic, Evidence-Based Practice/*organization & administration, *Public Policy, Policy Making, Politics

Citation Kebede M, Oxman AD, Glenton C, McKee M, Mills EJ, Pang T, Whitworth J. Blueprints for informed policy decisions: A review of laws and policies requiring routine evaluation. Report from Kunnskapssenteret no.16–2012. Oslo: Norwegian Knowledge Centre for the Health Services, 2012.

Norwegian Knowledge Centre for the Health Services summarizes and disseminates evidence concerning the effect of treatments, methods, and interventions in health services, in addition to monitoring health service quality. Our goal is to support good decision making in order to provide patients in Norway with the best possible care. The Centre is organized under The Norwegian Directorate for Health, but is scientifically and professionally independent. The Centre has no authority to develop health policy or responsibility to implement policies.

Acknowledgements This study was funded by the Norwegian Agency for Development Cooperation (Norad).

We thank all those who provided information about examples of laws and policies that require routine evaluation. We are especially grateful to: Anne Routhier, Senior Director and Suzanne Marshall, Senior Analyst, Centre of Excellence for Evaluation, Treasury Board of Canada, Canada; Ignacio Irarrázaval, Director of the Center for Public Policy, Pontificia Universidad Católica de Chile, Santiago, Chile; Diego Dorado, Director of the Public Policy Evaluation Bureau, National Planning Department, Bogota, Colombia; Edgar Adolfo Martinez Mendoza, Deputy Director General Coordination, National Council for the Evaluation of Social Development Policy, Mexico City, Mexico; and Grover (Russ) Whitehurst, Director of the Brown Center on Education Policy, The Brookings Institution , Washington, DC, USA.

The following people provided helpful feedback on the case reports: Amir Attaran, Institute of Population Health and Faculties of Law and Medicine, University of Ottawa, Canada; John Mayne, Advisor on Public Sector Performance, Canada; Cristian A. Herrera, Unit for Health Policy and Systems Research, School of Medicine, Pontificia Universidad Católica de Chile, Chile; Paula Darville, Thomas Leisewitz Velasco and María Hamuy, National Budget Bureau (DIPRES), Ministry of Finance, Chile; Amelia Fernandez, Department of Preventive Medicine and Carlos Gómez-Restrepo, Department of Clinical Epidemiology and Biostatistics, Pontificia Universidad Javeriana, Colombia; Francisco Becerra, Council on Health Research for Development, Mexico; Gonzalo Hernández Licona, Executive Secretary, CONEVAL, Mexico; Miguel Angel Gonzalez Block, Centre for Health Systems Research, National Institute of Public Health, Mexico; Octavio Gómez Dantés, Centre for Health Systems Research, National Institute of Public, Mexico; Beth Ann Bryan, Senior Education Advisor, Akin Gump, USA; Jon Baron, President, Coalition for Evidence-Based Policy, USA.

The following provided helpful advice regarding the protocol for this study, potential examples and a draft version of this report: Abu Bakar Suleiman, Amir Attaran, Andrew Podger, Andy Haines, Arild Bjørndal, Bertha Briceño, Bob Wells, Bocar Amadou Kouyate, Carmen Hooker Odom, David Bloom, David Fidler, Douglas William Bettcher, Francisco Becerra-Posada, Francisco Songane, Gemma Vestal, Haichao Lei, Haluk Soydan, Hassan Mshinda, Iain Chalmers, Judith Healy, Julia Littell, Maimunah A. Hamid, Manuel Dayrit,

Marie Gaarder, Mark Gibson, Mark Lipsey, Mark Petticrew, Nelson Sewankambo, Octavio Gómez Dantés, Pierre Ongolo-Zogo, Ruth Levine, Sara Bennett, Stella Bialous, Suwit Wibulpolprasert. (See Appendix 3 for a list of institutional affiliations.)

Contributors All of the authors reviewed and commented on the protocol for this study and the drafts of this report. Andy Oxman prepared a first draft of the protocol. Mekdes Kebede collected the data and conducted the analysis together with Andy Oxman and Claire Glenton. Mekdes Kebede, Claire Glenton and Andy Oxman prepared the first draft of this report. All of the authors contributed to responding to comments on the draft report.

Conflicts of interest None known.

Norwegian Knowledge Centre for the Health Services
Oslo, December 2012

Key messages

Countries are committed to improving the health and welfare of their populations. Yet, we found only five examples of laws and policies requiring routine evaluation of public programmes. This suggests that the majority of countries and international organisations may not be fulfilling their political and ethical obligations to use well designed evaluations of policies and programmes routinely to inform decisions about how best to use available resources to achieve societal goals. It is possible, however, that existing laws and policies do not adequately reflect the degree to which such actions are already being undertaken.

A number of important lessons can be drawn from the experiences discussed in this review, including:

- The enactment of laws and policies supporting routine use of evaluation can:
    - Capitalise on broad reforms (e.g. those focusing on accountability and transparency)
    - Build on existing laws and policies and existing evaluation experience
    - Be championed by a wide range of people, including auditors general, budget bureaus, multilateral organisations and donors, legislative branches of government, and heads of state
- Linking evaluation and monitoring objectives to other government initiatives and institutions can create synergies in budgetary processes and accountability and transparency, in ways that create an integrated rather than overly-regulated system
- Implementation of a monitoring and evaluation system goes hand in hand with administrative reforms. Such changes enable those responsible for monitoring and evaluation to respond to the information needs of decision-makers and to link monitoring and evaluation to decision-making
- It is important to focus clearly on assessing the performance of programmes in order to identify the core issues that need to be addressed in evaluations (e.g. effectiveness, efficiency and equity), and to make clear what type of evidence is wanted

Title:
Blueprints for informed policy decisions: A review of laws and policies requiring routine evaluation

---

Type of publication:
Review of laws and policies

Doesn't answer everything:
We were only able to identify five examples of laws and policies that made evaluation an expected element of decisions about public programmes rather than an optional extra. Given the small number of cases identified and limitations regarding how these have been evaluated it is not possible for us to draw firm conclusions.

Publisher:
Norwegian Knowledge Centre for the Health Services

Updated:
Last search for cases: December, 2010.

Peer review:
See acknowledgements

- It is advisable to have an entity with a clear mission to carry out independent, unbiased evaluations to a high standard. The entity should be insulated from the influence of political organisations or interest groups
- Monitoring and evaluation systems need to be supported by reliable and objective information which is continuously improved in order to give the system credibility
- A combination of flexibility and mandatory requirements is important
- Informing the public and engaging a wider spectrum of stakeholders in the design and interpretation of evaluation results will increase the probability that evaluation systems address questions that are important to intended beneficiaries, that the results of evaluations are used appropriately and, ultimately, that democratic principles are supported
- An appropriate legal framework and a well-designed and financed evaluation system can have important benefits
- There appears to be little risk of undesirable effects, although concerns have been raised about potential downsides, such as poor enforcement and bureaucratic implementation
- Laws and policies requiring routine evaluation should themselves be routinely evaluated

Executive summary

The problem

Substantial sums of money are invested each year in public programmes and policies, ranging from attempts to improve health, social welfare, education, and justice, to programmes related to agriculture, work, and technology. Little is known, however, about the effects of most attempts to improve lives in this way and whether public programmes are able to fulfil their primary objectives, such as enhancing health and welfare. What little is known is often not used to inform decisions.
Because public resources are limited, it is important to use them effectively, efficiently and equitably. This is essential in low- and middle-income countries faced with severe resource constraints and competing priorities. It is also essential in high-income countries where there are also limited resources and unmet needs, and the potential for waste is greater.
When making decisions about public programmes, good intentions and plausible theories alone are insufficient. Research evidence, values, political considerations, and judgements are all necessary for well informed decisions. However, decisions are often made without systematically or transparently accessing and appraising relevant research evidence and without an adequate evaluation of both the intended and unintended effects of programmes. We need to make better use of what we already know and to better evaluate the effects of what we do.
The problem of public programmes being affected by poorly informed decisions varies in scale from country to country, across international and non-governmental organisations, from sector to sector within countries, and across programmes within a sector. Similarly, the cause of the problem can also vary in scale, such as the availability of human and financial resources. Decisions about public programmes are sometimes well-informed by research evidence and are sometimes rigorously evaluated without explicit processes or criteria for deciding when to undertake an impact evaluation. However, across national settings and different sectors, relevant evaluations are frequently not used to inform decisions and the need to evaluate the effects of programmes is frequently not considered. A formal requirement to consider relevant research evidence and the need for evaluation routinely, systematically and transparently might help to ensure better use of research evidence, planning of evaluations, use of public resources and outcomes.

Policy options

Many initiatives exist that aim to improve the use of relevant evaluations to inform decisions about public programmes and decisions about when to evaluate the effects of such programmes. These include attempts to:

- Prioritise research and align it with the needs of countries
- Build the capacity to undertake evaluations
- Increase funds for evaluation
- Commission research to meet the needs of policymakers for better information
- Improve the quality of research syntheses and impact evaluations
- Make research evidence more accessible to policymakers (e.g. through the use of summaries of systematic reviews, clearing houses, and policy briefs)
- Build policymaker interest in evaluations and their capacity to use them
- Improve public understanding of research evidence and its role in informing decisions about public programmes.

Yet, relatively little attention has been paid to requirements for routine evaluation. We were able to identify few examples of such requirements (Box 1). All of these appear to improve the use and conduct of evaluations and none appear to have important undesirable effects. However, given the small number of cases and the limitations associated with how these have been evaluated, it is not possible to draw firm conclusions. Details of how we identified and reviewed these five cases are described in our full report.

Box 1. Examples of requirements for routine evaluation

Canadian Policy on Evaluation
The Treasury Board of Canada Secretariat is the central agency responsible for providing leadership for evaluation across the Canadian federal government, and gives advice and guidance in the conduct, use and advancement of evaluation practices. Deputy heads of department are responsible for establishing a robust, neutral evaluation function in their departments and for ensuring that their department adheres to the Policy on Evaluation and its supporting directive and standard.

Chilean Budget Bureau's Evaluation System
The Ministry of Finance must formulate one or more decrees specifying which programmes or projects will be evaluated each year. The Evaluation Programme forms part of the Management Control System and is located in the National Budget Bureau (DIPRES) at the Ministry of Finance.

Colombian Monitoring and Evaluation System
The National Planning Department was given responsibility for creating the National System for Monitoring and Evaluation (SINERGIA) and for reporting annually to the National Council for Economic and Social Policy (a policy committee headed by Colombia's President) on the evaluation findings. A National Planning Department resolution assigned responsibility for self-evaluation to all agencies in the executive branch of the government. The Directorate for Evaluation of Public Policies, a unit established within the National Planning Department is the technical secretariat of SINERGIA.

Mexican Laws for Social Development and Financial Responsibility, and General Guidelines for the Evaluation of Federal Programmes
The National Council for the Evaluation of Social Development Policies (CONEVAL) has the power – based on the General Law for Social Development – to regulate and coordinate the evaluation of social development policies and programmes and to assess periodically the compliance of programmes with their social objectives. The Secretariat of Finance and Public Credit and the Secretariat of Public Service, together provide a system of performance evaluation – based on the Federal Budget and Financial Responsibility Law – to evaluate the efficiency, economy, effectiveness and social impact of public expenditure. The Secretariat of Public Service evaluates the performance and results of the relevant institutions. All federal secretariats and agencies are required to adhere to the evaluation guidelines and must use the prescribed monitoring and evaluation instruments.

USA Evaluation of Educational Programmes

There is no overarching body responsible for programme evaluation. A case-by-case assessment is made for each programme to determine the specific manner in which it will be evaluated. For several years, two offices in the Department of Education have been responsible for programme and policy evaluation. The Policy and Program Studies Service based in the Office of Planning, Evaluation, and Policy Development, advises the Secretary on policy development and review, strategic planning, performance measurement, and evaluation. The Institute of Education Sciences (IES) is the research arm of the Department. The IES is charged with producing rigorous evidence on which to ground education practice and policy, with programme evaluation being undertaken primarily by the National Center for Education Evaluation and Regional Assistance.

The five cases identified illustrate a variety of options for designing and implementing programme requirements (Table 1).

Table 1. Options for how to design and implement requirements for routine evaluation

Considerations	Options
Enactment of laws and policies	The enactment of requirements for routine evaluation can be precipitated by a variety of events (such as the election of a new government), can have a variety of motivations (such as improving expenditure decisions or transparency), can be championed by a range of advocates (such as an auditor general or a president), and can build on earlier laws and policies and on experience with evaluation.
Scope of laws and policies	Requirements can apply across sectors or within a sector. They can also apply to the use of research evidence to inform decisions about programmes, decisions about when and how to undertake evaluations, or both. However, the five examples that we identified and reviewed only focused on decisions about when and how to undertake evaluations.
Responsibility for enforcing laws and policies	The primary responsibility for enforcing laws and policies can be vested in a treasury department (linked to budgetary processes), in a planning department (linked to planning processes), in an independent organisation, spread across departments and agencies, or a combination of these.
How laws and policies are enforced	Strategies for enforcing the laws and policies include: having identifiable people or organisations responsible and accountable for evaluation, monitoring compliance, and taking corrective actions; and the real or perceived power to withdraw funding when there is a lack of adherence. There may also be mechanisms to ensure compliance that form part of the general structure of government (e.g. clear, understood and accepted responsibilities and accountability) or other legislation or policies (e.g. incentives that can be used as an incentive for civil servants to undertake evaluation).
Decisions about which programmes to evaluate	Approaches to deciding which programmes need evaluation include requiring evaluation (not necessarily impact evaluation) for all programmes while allowing flexibility in deciding on the approaches to be used or providing a structured process for deciding which programmes to evaluate. Structured processes can engage a variety of stakeholders and use different criteria and processes adapted to specific contexts.
Who undertakes evaluations	Evaluations can be commissioned, can be undertaken in-house, or both.
Specification of methods used in evaluations	The methods used in specific evaluations can be determined by the people responsible for undertaking the evaluation, by a central entity responsible for evaluations, by the department responsible for the programme being evaluated, or by a combination of these approaches.
Funding for evaluations	Funding mechanisms can include the allocation of core funding to entities responsible for evaluation, earmarked funds for evaluation linked to programmes, external funding, and requirements for departments to pay for evaluations from their own budget.
Enforcement of recommendations derived from evaluations	Ways to ensure that the evaluation results are used include: assuring the relevance and legitimacy of evaluations; designing evaluations to generate not only impact assessment but hypotheses about ways to improve programmes; framing conclusions in a way that will not alienate those responsible for the programmes; forums within the legislative and executive branches and within civil society; the joint drafting of institutional commitments by the organisation responsible for the evaluation and the organisation responsible for the programme; assigning responsibility to the senior civil servant in each department; a follow-up report on the aspects of public programmes that can be improved; an evaluation report on social development policy that establishes recommendations addressed to different decision makers; a performance evaluation system that provides information for budgetary decision-making; and the monitoring of compliance with commitments.

Considerations	Options
Transparency and independence	Requirements for transparency vary in relation to different types of decisions, including: which programmes are evaluated, who will undertake evaluations, what methods are used in evaluations, how the results of evaluations are reported and disseminated, and how evaluations are used. Similarly, requirements for independence can vary in relation to who pays for evaluations, decisions about which programmes are evaluated, who undertakes evaluations, decisions about the terms of reference for evaluations, decisions about the methods that are used in evaluations, reporting and interpreting the results of evaluations, peer review of evaluation reports, and decisions about how the results of evaluations are used.
Evaluation of laws and policies	We did not find any evaluations that compared outcomes of any kind in settings with and without requirements for evaluation. Assessments of existing requirements have been undertaken by external groups in Chile, Colombia and Mexico and, to some extent, internally in all five countries. While these assessments have largely been positive, a number of challenges have been identified, including concerns about human and financial capacity (and consequently only a small proportion of programmes being evaluated). The absence of a clear link between evaluation and planning and budgeting processes, including decisions about modifying or discontinuing programmes, has also been identified as a concern.

The enactment of requirements for routine evaluation has been prompted by various factors. In four of the cases we identified (Chile, Colombia, Mexico, and the United States of America), laws and policies were initiated by new governments as part of a broader set of reforms focusing on or motivated by a need to improve the effectiveness of state policies and programmes, expenditure decisions, and public management. Additional concerns included the perceived need to improve systems of evaluation because of concerns about corruption, and to counter a perceived lack of objectivity, technical rigour, transparency and accountability. Establishing a body outside government, which would be devoted to evaluating programmes and focused on results-based management, was therefore perceived as necessary. Requirements for routine evaluation were championed or supported by a range of different stakeholders in these cases, including auditors general, budget bureaux, heads of state, parliaments or legislative branches of government, individual Members of Congress, as well as by multilaterals and donors. The enactments built upon earlier laws and policies, a culture of evaluation, and past evaluation experience.

The advantages and disadvantages of using an intersectoral versus a sectoral scope may be affected by the size of a particular country and sector. In the United States of America (USA), for example, more resources are used for evaluation within the education sector alone (US$70 million annually) than across sectors in Chile, Colombia and Mexico (ranging from US$2.5 to US$8 million annually per country). Trade-offs may need to be made between an increased potential for independence and efficiency afforded by being outside a sector versus the increased potential for ownership and communication when inside a sector. Canada has attempted to capitalise on the advantages of using both an intersectoral *and* sectoral approach by applying the Treasury Board's Policy on Evaluation to *all* government spending without precluding other departments from having their own specific policies. Health Canada, for example, is therefore allowed to use its own policy to make evaluations more specific to the health sector. In addition, every department is required to have an evaluation function; a central entity for evaluations and the use of guidelines helps to ensure consistent quality standards.

The USA was the only example, of the five cases we reviewed, in which recommendations are not included in evaluation reports. While reports may include recom-

mendations for further research, they do not include recommendations for policy decisions. The rationale for this approach is that when recommendations are made, this introduces subjective values and political standpoints. While informants from the other countries acknowledged the importance of this concern, they believed that there was still a need for recommendations within their own national context.

We identified the following key strengths in the five examples of requirements for routine evaluation: the extensive use of information in budget- and decision-making, the ability to monitor progress towards political goals, the active participation of key stakeholders in monitoring and evaluation activities, independent evaluation and appropriate levels of financial support, a strong monitoring and evaluation system, and improvements in research capacity and quality. Key weaknesses in one or more of the cases were:

- A failure to adequately clarify roles and responsibilities
- A lack of evaluation, oversight and accountability functions
- A lack of comprehensive coverage and inappropriate discretion in deciding which programmes to evaluate
- Restrictions on how contracts for evaluations are awarded
- Problems with the availability and quality of data
- The absence of clear links between evaluation and planning and budgetary processes
- The low utilisation of the results of evaluations and nonbinding recommendations
- A failure to build capacity and disseminate results to subnational authorities
- A lack of evidence-based programmes despite official requirements to have them

Implementation considerations

Challenges to implementing requirements for routine evaluation include:

- A lack of skilled people to manage the processes or undertake evaluations
- Inadequate financing
- A lack of routinely collected data or the means to collect reliable data for evaluations
- A lack of awareness of the benefits of evaluation
- Too much discretion being used during evaluations
- Poorly defined programmes (i.e. what the focus of an evaluation is)
- Complex legal frameworks caused by multiple pieces of legislation and policies
- Procurement legislation that makes it difficult to commission evaluations

Strategies to address these problems include:

- Linking evaluation and monitoring objectives to other government initiatives and institutions in order to create synergies in budgetary processes, accountability and transparency, and thereby creating integrated systems that are not overly regulated

- Administrative reforms which enable those responsible for monitoring and evaluation to respond to the information needs of decision-makers and to link monitoring and evaluation to decision-making
- A clear focus on assessing programme performance and identifying the core issues that should be addressed in evaluations (e.g. effectiveness, efficiency and equity) and a clear specification of what type of evidence is needed
- Having an entity with a clear mission to carry out independent, unbiased evaluation to a high standard. The entity should be insulated from the influence of political organisations or interest groups
- Supporting monitoring and evaluation through the use of reliable and objective information which is continuously improved, thereby giving credibility to evaluations
- Applying both mandatory and flexible requirements
- Informing the public and engaging a wide spectrum of stakeholders in the design and interpretation of evaluation results to ensure that evaluation systems address questions that are important to intended beneficiaries, that the results of evaluations are used appropriately and, ultimately, support democratic principles

Next steps

A first step should be the assessment of the size of the problem (poorly informed decisions about public programmes) and its causes within the specific context. In most instances, there are no formal demands for routinely considering relevant research evidence systematically and transparently, or for using evaluations when making decisions about public programmes. However, a consideration of such requirements is warranted. The design and implementation of requirements for routine evaluation can be informed by the experience summarised in this report and by related experience, including findings from institutionalising evaluations, health impact assessments, environmental impact assessments, health technology assessments, and regulatory impact assessments. Arguments against such requirements should also be considered. If there is inadequate evaluation capacity, poor implementation, or an under-developed culture of evaluation there is a risk that evaluations may simply become bureaucratic requirements that need to be 'checked off'. Laws and policies requiring routine evaluation should themselves be routinely evaluated.

Hovedfunn (norsk)

En av statens viktigste oppgaver er å bidra til forbedringer i befolkningens helse og velferd. Til tross for det: I vår gjennomgang av internasjonale erfaringer fant vi kun fem eksempler på lover og regler med pålegg om rutinemessig evaluering av offentlige reformer og programmer. Dette kan tyde på at de fleste land og internasjonale organisasjoner ikke oppfyller sin politiske og etiske forpliktelse til rutinemessig å evaluere reformer og programmer – som grunnlag for beslutninger om hvordan ressurser best bør brukes for å oppnå vedtatte mål. Det er imidlertid mulig at eksisterende lover og regler ikke i tilstrekkelig grad gjenspeiler hvilken grad slike evalueringer faktisk blir gjennomført.

En rekke viktige lærdommer kan trekkes fra denne gjennomgangen av erfaringer med lovpålagt evaluering, blant annet:

- Vedtak av lover og regler for rutinemessig bruk av evaluering kan:
    - Dra nytte av brede reformer (f.eks reformer med fokus på plassering av ansvar og større åpenhet)
    - Bygge på eksisterende lover og regler og tidligere evalueringserfaringer
    - Støttes av et bredt spekter aktører, blant annet riksrevisjon, finansdepartement, multilaterale organisasjoner og givere, lovgivende forsamlinger, og statsoverhoder
- Kobling av evaluering og monitorering til andre offentlige initiativ og institusjoner kan skape synergier i budsjettprosesser, og bidra til større grad av ansvarlighet og åpenhet, i et integrert og ikke overregulert system
- Iverksetting av et system for monitorering og evaluering går hånd i hånd med administrative reformer: Det gjør det mulig for dem med ansvar for monitorering og evaluering å møte informasjonsbehovet til beslutningstakerne, og å knytte monitorering og evaluering til beslutningsprosessene
- Det er viktig å være opptatt av hva programmene har ført til, for å identifisere de viktigste spørsmålene som må belyses i evalueringen (f.eks virkningene av programmet, og konsekvenser med tanke på sosial ulikhet), og for å gjøre det klart hva slags type dokumentasjon som trengs

Tittel:
Internasjonale erfaringer med lovpålagte krav til rutinemessig evaluering

---

Publikasjonstype:

Oversikt av lover og regler

Svarer ikke på alt:

Vi fant kun fem eksempler på lover og regler med pålegg om rutinemessig evaluering av offentlige reformer og programmer, i vår gjennomgang av erfaringer internasjonalt. Gitt det lave antallet tilfeller av lovpålagt evaluering som vi identifiserte, samt svakheter ved hvordan disse i seg selv er blitt evaluert, er det ikke mulig for oss å trekke bastante konklusjoner om nytten av slike lover og regler. -

---

Hvem står bak denne rapporten?

Forfatterne har skrevet rapporten på oppdrag fra Verdens helseorganisasjon (WHO) sin Advisory Committee on Health Research.

Når ble søket utført?

Søk etter caser ble avsluttet desember 2010.

Fagfeller:

Se “Acknowledgements”.

- Én enhet bør få som sin oppgave å utføre uavhengige, objektive evalueringer av høy kvalitet. Enheten bør skjermes for påvirkning fra politiske grupperinger og interesseorganisasjoner
- En forutsetning for et troverdig monitorerings- og evalueringssystem er at det har tilgang til pålitelig og objektiv informasjon
- Det er viktig å både tillate fleksibilitet og stille krav
- Å informere publikum og engasjere et bredt spekter av interessenter i utforming av evalueringer og fortolkning av evalueringsresultater øker sannsynligheten for at spørsmålene som er viktige for disse gruppene inngår i evalueringene, at resultatene av evalueringene brukes på en hensiktsmessig måte og, i ytterste konsekvens, en styrking av demokratiske prinsipper
- Et fornuftig juridisk rammeverk og et godt utformet og finansiert evalueringssystem kan gi betydelige gevinst
- Risikoen for uønskede konsekvenser synes liten, men mulige problemer er for eksempel dårlig håndhevelse og for byråkratisk gjennomføring av regelverket
- Lover og regler som innebærer pålegg om rutinemessig evaluering bør selv evalueres rutinemessig

Nasjonalt kunnskapssenter for helsetjenesten fremskaffer og formidler kunnskap om effekt av metoder, virkemidler og tiltak og om kvalitet innen alle deler av helsetjenesten. Målet er å bidra til gode beslutninger slik at brukerne får best mulig helsetjenester. Kunnskapssenteret er formelt et forvaltningsorgan under Helsedirektoratet, men har ikke myndighetsfunksjoner og kan ikke instrueres i faglige spørsmål.

Nasjonalt kunnskapssenter for helsetjenesten
PB 7004 St. Olavs plassN-0130 Oslo, Norway
Telefon: +47 23 25 50 00
E-mail: post@kunnskapssenteret.no
Hele rapporten (pdf): www.kunnskapssenteret.no/Publikasjoner

Sammendrag (norsk)

Bakgrunn

Myndighetene er moralsk forpliktet til å fatte beslutninger på solid grunnlag, og rutinemessig evaluering av virkningene av offentlige reformer og programmer er nødvendig for å oppfylle denne forpliktelsen. En vurdering av potensielle virkninger bør utføres før programmer begynner, og det kan være behov for å evaluere de faktiske virkninger etter at tiltakene er iverksatt.

Mål

Formålet med denne gjennomgangen var å identifisere og lære av eksempler på lover og regler som pålegger rutinemessig evaluering av offentlige programmer.

Metoder

For å identifisere relevante eksempler benyttet vi oss av personlige og sosiale nettverk for å identifisere kontakter i så mange land og internasjonale organisasjoner som mulig. Vi søkte også etter aktuell litteratur i flere databaser. Eksempler på lovpålagt evaluering ble inkludert dersom følgende kriterier var oppfylt: lover eller retningslinjer som inneholder krav om en eksplisitt prosess eller klare kriterier for å avgjøre når evaluering skal foretas, eller lover eller regler som pålegger evaluering av alle programmer på et gitt tematisk område.

Resultater

Vi identifiserte fem tilfeller av lovpålagt evaluering som oppfylte våre inklusjonskriterier. Disse var fra Canada, Chile, Colombia, Mexico og USA. I ingen av disse eksemplene forelå det krav om rutinemessig bruk av systematiske oversikter av relevant forskning før iverksetting av nye programmer. Det var heller ingen av dem som fokuserte spesifikt på evaluering av virkninger (effektevaluering). Disse lovene og reglene dro fordel av pågående brede reformer som fokuserte på plassering av ansvar og økt åpenhet, baserte seg på eksisterende lov- og regelverk, og hadde et bredt spekter av støttespillere. Viktige styrker i lover og regler som påla evaluering

var blant annet utstrakt bruk av informasjon i budsjetterings- og beslutningsprosesser, monitorering av framdrift mot politisk vedtatte mål, at sentrale aktører tok aktiv del i monitorerings- og evalueringsaktiviteter, uavhengig evaluering, og adekvat økonomisk støtte. Viktige svakheter som ble rapportert fra ett eller flere av landene var behov for større tydelighet når det gjaldt å avklare roller og ansvar, for få gjennomførte evalueringer, for stor grad av skjønnsmessig vurdering ved avgjørelser om hvilke programmer som skal evalueres, for dårlig tilgjengelighet og kvalitet på data, svake koplinger mellom evaluering og planleggings- og budsjettprosesser, og at resultatene fra evalueringene ble benyttet i for liten grad.

Konklusjoner

Mange land stiller krav om rutinemessig evaluering av offentlige programmer, for eksempel konsekvensutredninger og medisinske metodevurderinger (HTA). Men vi klarte bare å identifisere fem eksempler på lover og regler der evaluering inngikk som en integrert del av beslutninger om å iverksette politiske reformer og programmer. Selv om noen land rutinemessig bruker systematiske oversikter (for eksempel medisinsk metodevurdering), inngikk ikke dette i noen av eksemplene som inngår i vår gjennomgang. Eksemplene vi har identifisert kan tjene som utgangspunkt for planlegging av systemer som gjør det mulig for myndigheter å oppfylle sin politiske og etiske forpliktelse til å foreta velinformerte beslutninger. Detaljert planlegging er viktig for å utforme effektive systemer. Men, det er mennesker som utfører evalueringer og mennesker som avgjør om resultatene fra en evaluering brukes på en fornuftig måte, og om systemene fungerer. Selv om disse eksemplene på lovpålagt evaluering er tilpasset en bestemt kontekst, er det likevel mye å lære fra erfaringene som er gjort i disse landene. De fem eksemplene viser at et fornuftig juridisk rammeverk og et godt utformet evalueringssystem kan gi betydelige gevinster, og faren for uheldige virkninger ser ut til å være liten. Gitt det lave antallet tilfeller av lovpålagt evaluering som vi identifiserte, samt svakheter ved hvordan disse i seg selv er blitt evaluert, er det ikke mulig for oss å trekke bastante konklusjoner om nytten av slike lover og regler. Det samme gjelder tross alt her for som for andre vedtatte reformer og programmer: Det er behov for solide evalueringer – rutinemessig.

Table of contents

Abstract

Background: Governments have ethical obligations to make well-informed decisions and routine evaluation of the impacts of public programmes is essential to fulfil these obligations. An assessment of potential impacts should be undertaken before programmes begin, and an evaluation of impacts may be needed after programmes have been launched.

Objective: The objective of this review was to identify and learn from examples of laws and policies requiring routine evaluation of public programmes.

Methods: To identify relevant examples we used personal and social networking to identify contacts in as many countries and international organisations as possible. We also searched the literature using multiple databases. Examples were included if they met the following inclusion criteria: laws or policies requiring an explicit process or criteria for deciding when to undertake an evaluation, or laws or policies requiring the evaluation of all programmes within a given area.

Results: We identified five cases of laws and policies that met our inclusion criteria. The cases were from Canada, Chile, Colombia, Mexico, and the USA. None required routine use of systematic reviews of relevant research prior to launching new programmes or focused specifically on impact evaluations. The enactment of these laws and policies capitalised on broad reforms that focused on accountability and transparency, built on existing laws and policies, and had been championed by a wide range of advocates. Key strengths of the laws and policies included the extensive use of information in budgetary processes and decision-making, the monitoring of progress towards political goals, the active participation of key stakeholders in monitoring and evaluation activities, independent evaluation, and the provision of appropriate levels of financial support. Key weaknesses identified in one or more of the five countries included: the need to further clarify roles and responsibilities, a lack of comprehensive coverage, too much discretion being given when deciding which programmes to evaluate, problems with the availability and quality of data, the absence of clear links between evaluation and planning and budgetary processes, and a low utilisation of the evaluation results.

Conclusions: Precedents for laws and policies requiring routine evaluation of public programmes include environmental impact assessments and mandatory health technology assessments. We were only able to identify five examples of laws and policies that made evaluation an *expected* element of decisions about public programmes rather than an optional extra. Although some countries routinely use systematic reviews (such as health technology assessments), none of the laws and policies that met our inclusion criteria required this. Each example serves only as a blueprint for systems that help to enable and ensure that governments fulfil their political and ethical obligation to make well-informed decisions in a specific context. Blueprints are essential when designing effective systems. But it is people who carry out evaluations and people who are essential to use the results appropriately and

ensure that these systems work. Although each set of laws and policies is a blueprint which has been adapted to a specific context, a number of lessons can be learned from these experiences. These five cases illustrate that an appropriate legal framework and a well-designed evaluation system can have important benefits, and we found little apparent risk of undesirable effects. Given the small number of cases identified and limitations regarding how these have been evaluated it is not possible for us to draw firm conclusions. As with any other public policy or programme, routine evaluation of such legal frameworks and systems is needed.

Background

Rigorous evidence is often lacking about the impacts of public investment in programmes to improve health, social welfare, education, and justice, despite the fact that these and other programmes have important impacts on people's lives. Even when evidence is available, it is often not used to inform decisions. Impact evaluations can help to counter such problems and ensure that programmes are effective, that they do not have adverse effects, and that money is well spent. The results of impact evaluations can inform decisions about whether to continue, expand, modify or discontinue programmes.

Three of the authors of this report have previously argued that governments should recognise – through legislation – their ethical obligations to make well-informed decisions and evaluate the impacts of their programmes [1]. Decisions about whether to conduct impact evaluations are often made on an ad hoc basis. But such obligations demand a requirement for the more systematic and transparent use of research evidence to assess the likely effects of public programmes before they are launched, and the better use of well-designed impact evaluations after programmes have begun. Mandating an explicit process or criteria for deciding when to undertake an impact evaluation or requiring routine evaluation of all programmes within a given area are two ways in which this might be achieved.

To facilitate the drafting, enactment and implementation of such legislation, we proposed that a framework be developed by the World Health Organization (WHO) to support governments that are seeking to make formal commitments – "mandatory impact evaluations" – to improve the use of research evidence. Deciding whether to develop a framework and, if so, how to design it is a process that should be informed by a review of relevant international experience. This study was therefore conducted with the aim of addressing this need.

In this report we refer to "laws and policies that require routine evaluation" rather than to "mandatory impact evaluation" to clearly avoid the implication that *all* programmes require impact evaluation. Programmes are sometimes evaluated without explicit processes or criteria for deciding when to undertake an impact evaluation. Frequently, however, relevant evaluations are not used to inform decisions and the need for impact evaluations is not considered. More formal requirements might help to ensure a better use of evaluations and better evaluation planning.

Several previous papers have examined what is known as the "institutionalisation of evaluation" [2-10]. Appendix 1 provides a summary of these papers and their key

messages are presented in Box 2. The institutionalisation of evaluation can be defined as "a process of channelling isolated and spontaneous program evaluation efforts into more formal and systematic approaches" [2]. Institutionalisation does not necessarily entail requirements for routine evaluation. On the other hand, requirements for routine evaluation do not necessarily entail the establishment of an entity with a clear mission to carry out impact evaluations. However, to be useful and meaningful, requirements for routine evaluation *must* include a legal mandate to conduct evaluations [2], allow for an organisation to commission or undertake them, and ensure that there is a system to link evaluation to decision-making. The way in which these can be achieved will vary according to the characteristics of government in different countries.

Box 2. Key messages from reviews of the institutionalisation of evaluation

- Locating responsibility for evaluation within the right institution and ensuring linkages to budget processes might be important for the success of institutionalisation.
    - The independence of the institution responsible for evaluation might lead to greater credibility and acceptability.
    - On the other hand, locating such responsibilities outside of government might limit influence on planning, implementation, and budget processes.
- Contracting out evaluations to external evaluators might help ensure the impartiality of assessments.
- Other factors that might be important to the success of institutionalisation include:
    - A system of democracy in which the voices of the opposition, strong stakeholders (such as the American Congress or a Ministry of Finance), and champions of evaluation are heard
    - A unique and broad legal mandate and laws to ensure access to public information and transparency
    - Institutional and human capacity
    - Systematising evaluation activities to make them regular rather than ad hoc activities
    - Involvement of intended users in identifying what programmes will be evaluated
    - Appropriate timing of evaluations

In this review, we examine considerations regarding the development, scope, characteristics and impacts of such legislation in detail. We focus specifically on legislation and policies requiring routine evaluation,[1] including those that establish explicit obligations to use evaluations to inform decisions prior to launching new programmes, or to undertake evaluations after programmes are launched.

Some types of impact assessment are mandatory in many countries around the world, including environmental impact assessments and regulatory impact assessments. Other types of impact assessments, such as health impact assessments, are generally not mandatory [11-30]. An overview of these different types of impact as-

[1] As noted above, we initially used the term "mandatory impact evaluation" in this research. However, this led to a misunderstanding that our focus was on a requirement for *every* programme and policy to be evaluated, whereas our focus is on *explicit* requirements for a systematic and transparent process for making decisions about the need for evaluation. In addition, our focus is not restricted to requirements for impact evaluations specifically. For these reasons we have referred to "requirements for routine evaluation" instead of "mandatory impact evaluation".

sessment is presented in Appendix 2. The focus of this review is on the experience and precedents that these different assessments provide in terms of requirements for routine evaluation of the effects of public programmes to improve health, social welfare, education or justice. Key lessons from reviews of these different types of impact assessment are summarised in Box 3.

Box 3. Key messages from reviews of health impact assessment, environmental impact assessment, health technology assessment, and regulatory impact assessment

Factors that might be important for the success of different types of impact assessment include:

- Organisational, statutory and policy commitments, including support from the highest levels of government (e.g. through laws or decrees)
- The integration of assessments into budgetary processes
- The appropriate timing of assessments
- Knowledge of the policy-making environment by those conducting assessments
- Realistic recommendations
- A mechanism through which policymakers are held accountable for using the results of assessments

Methods

Identification of examples of requirements for routine evaluation

We established an advisory group which consisted of 38 individuals from 20 countries and represented a broad range of relevant backgrounds and perspectives (Appendix 3). Members were asked to review the project proposal, identify relevant examples of mandatory impact evaluation, and to identify additional people who might be able to offer further examples. A similar request was sent to those identified by the advisory group, to personal contacts, and to any additional people identified via these networking processes.

Contact was also made with relevant organisations, including United Nations agencies, international development agencies, and multilateral development banks. Members of relevant organisations were reached via discussion lists or newsletters, including the Cochrane Collaboration, the Grade Working Group, the European Science Advisory Network for Health, the Alliance for Health Policy and Systems Research, the International Network of Agencies for Health Technology Assessment, the International Clinical Epidemiology Network, the Guidelines International Network, the Health Impact Assessment Network, the Evidence-Informed Policy Network, the African Evaluation Association, the European Evaluation Association, and the International Association for Impact Assessment.

Relevant publications and grey literature were searched to identify examples and additional contacts, to find relevant background documents describing experiences related to the mandatory evaluation of public programme impacts, and to locate reviews of other types of impact assessment (Appendix 2). In addition to using Google Scholar, PubMed, the Global Legal Information Network, and the International Digest of Health Legislation to identify relevant literature, we also searched the websites of relevant organisations, including those of regional evaluation associations. Examples include United Nations evaluation groups, the International Network for Impact Assessment, and the International Network of Agencies for Health Technology Assessment.

We included legislation, regulations or policies (laws and policies) that:

- Mandate the systematic and transparent use of evaluations prior to launching a programme, or evaluations after programmes are launched
- Mandate an explicit process or criteria for deciding when to undertake an evaluation or mandate the evaluation of all programmes within a given area

- Establish only limited requirements, apply only to a limited range of programmes, or apply to areas other than health, social welfare, education or justice, but may nonetheless provide important lessons
- Are national, subnational (e.g. for states or provinces), or international (e.g. for a United Nations agency or multilateral development bank)

We also considered laws and policies that include requirements for impact evaluations but do not focus specifically on impact evaluation. Requirements for the evaluation of a single specific programme were excluded (those embedded, for example, in legislation to establish a programme). We also excluded legislation that establishes funding or institutions for evaluation without also requiring evaluations.

Development and distribution of a questionnaire

We designed a questionnaire (Appendix 4) that would help us to learn as much as possible from each identified case, and how best to inform decisions about enacting, designing, strengthening and evaluating mandatory impact evaluation. The questionnaire included questions about:

- When, why and how the laws and policies were enacted
- The scope of the laws and policies
- The characteristics of the laws and policies, including who is responsible for implementation, how the laws and policies are enforced, how decisions are made about which programmes to evaluate, who does the evaluations, how decisions are made about which evaluation methods to use, and how the use of the results of the evaluations is ensured
- The effects of the laws and policies
- Advocates and critics of the laws and policies
- Examples of successes and failures of the laws and policies in achieving their desired goals
- The strengths and weaknesses of the laws and policies in relation to the topics above

A draft version of the questionnaire was circulated to members of the advisory group for their comments and changes were made based on the feedback received. The revised pilot questionnaire was then emailed to three key informants each of whom completed it for a case. No further changes were required based on this pilot test.

For each case that met our inclusion criteria, we identified a key informant to whom we sent the questionnaire by email.

Preparing case summaries

For each example of a routine evaluation requirement, one of us prepared a summary based on the questionnaire responses, additional data (including the text of the laws and policies), and other reports describing the legislation and evaluation systems or evaluations of the legislation and evaluation systems (Appendix 5). New

information and perspectives that we had not directly elicited through the questionnaire were included if they were relevant to enacting, designing, strengthening or evaluating requirements for routine evaluation. Each summary followed the structure of the questionnaire, but changes were made to improve the flow of the text if needed.

Collecting additional information and checking the case summaries

Two of the authors checked each case summary and noted any details that were unclear or missing. Instances in which details mentioned in one case might also be relevant to others were also noted. To ensure the accuracy of the summaries, such issues were used to develop further questions. These, together with additional follow-up questions, were sent to respondents to ensure the accuracy of the summaries. Responses to the questions were obtained by telephone interviews with the key informants. In addition, at least two additional informants who had a different perspective for each case were identified and asked to cross-check the veracity of our summary. A draft of the full report was sent to them with a request for comments on the specific case report.

Interviews were conducted in English, but one was undertaken with the help of a translator. With the consent of the participants, four of the five interviews were recorded.

Data analysis

A thematic analysis of the case summaries was undertaken using elements drawn from a framework analysis approach [31]. Three of the report authors independently read and re-read each summary, identified key themes and coded data. Following this, these authors met to compare, discuss and agree upon a common thematic framework and prepared and circulated a draft report to the advisory group and all the respondents. Finally, we compiled all the comments received and responded to each one.

Results

In total, 39 potential cases were identified, five of which met our inclusion criteria. These cases are summarised in Table 2. Appendix 5 provides more detailed descriptions of the cases included, while an overview of the key features of the included laws and policies is provided in Table 3. The 34 cases that were excluded (and the reasons for excluding them) are listed in Appendix 6. For the sake of brevity, we have reported opinions elicited from the key informants without stating each time that these were expressed by the informants. We have not included our own inferences in the results.

Table 2. Summaries of the five included cases

Country	Laws and policies	Scope	Organisational responsibility
Canada	Federal Accountability Act (2006) Policy on Evaluation (2009)	Intersectoral	The Treasury Board of Canada Secretariat is the central agency responsible for providing leadership for evaluation across the Canadian federal government, giving advice and guidance in the conduct, use and advancement of evaluation practices. Deputy heads of departments are responsible for establishing a robust, neutral evaluation function in their department and ensuring that their department adheres to the Policy on Evaluation and its supporting directive and standards.
Chile	Law 19896 (2003) Ministry of Finance Decree 1177 (2003) Instructions for the Execution of the Public Sector Budget Law (2004)	Intersectoral	The Ministry of Finance must formulate one or more decrees annually specifying which programmes or projects will be evaluated during the year. The Evaluation Programme forms part of the Management Control System, located in the National Budget Bureau (DIPRES), in the Ministry of Finance.
Colombia	Constitution (1991) Resolution nr 63 (1994) Ley organica 152 de 1994 Conpes 3294 de 1994 Ley 1152 de 2007	Intersectoral	In 1994 the National Planning Department was given responsibility for creating the National System for Monitoring and Evaluation (SINERGIA) and for reporting annually to the National Council for Economic and Social Policy (a policy committee headed by the president) on the results of the evaluation system. A National Planning Department resolution in 1994 assigned responsibility for self-evaluation to all agencies in the executive branch of government. The Directorate for Evaluation of Public Policies, a unit established within the National Planning Department, is the technical secretariat of SINERGIA.
Mexico	Federal Budget and Financial Responsibility Law (2001) General law of social development (2004) Guidelines for evaluation (2007) Ley 14450 (2011)	Intersectoral	The National Council for the Evaluation of Social Development Policies (CONEVAL), has the power – based on the General Law for Social Development – to regulate and coordinate the evaluation of social development policies and programmes and the periodic assessment of whether programmes comply with their social objectives. The Secretariat of Finance and Public Credit and the Secretariat of Public Service, based on the Federal Budget and Financial Responsibility Law, form a performance evaluation system to identify the efficiency, economy, effectiveness and social impact of public expenditure, and, in the case of the Secretariat of Public Service, evaluating the performance and results of institutions. All federal secretariats and agencies are required to adhere to the guidelines for evaluation and must use the prescribed monitoring and evaluation instruments.
USA	The No Child Left Behind Act (2001) Education Sciences Reform ACT (2002) The Individuals with Disabilities Education Improvement Act (2004)	Education	There is no overarching body responsible for evaluation. A case-by-case assessment is made for each programme to determine the specific manner in which it is evaluated. For several years, two offices in the Department of Education have been responsible for programme and policy evaluation. The Policy and Program Studies Service, in the Office of Planning, Evaluation, and Policy Development, advises the Secretary on policy development and review, strategic planning, performance measurement, and evaluation. The Institute of Education Sciences (IES), established in 2002 (and replacing the Office of Educational Research and Improvement), is the research arm of the Department of Education. The IES is charged with producing rigorous evidence with which to ground education practice and policy. Programme evaluation is housed primarily in the National Center for Education Evaluation and Regional Assistance.

Table 3. Overview of the included laws and policies

	Canada	Chile	Colombia	Mexico	USA
Laws and policies	• Federal Accountability Act (2006) • Policy on Evaluation (2009)	• Law 19896 (2003) • Ministry of Finance Decree 1177 (2003) • Instructions for the Execution of the Public Sector Budget Law (2004)	• Constitution (1991) • Resolution nr 63 (1994) • Ley organica 152 de 1994 • Conpes 3294 de 1994 • Ley 1152 de 2007	• Federal Budget and Financial Responsibility Law (2001) • General Law of Social Development (2004) • Guidelines for Evaluation (2007) • Ley 14450 (2011)	• The No Child Left Behind Act (2001) • Education Sciences Reform Act (2002) • The Individuals with Disabilities Education Improvement Act (2004)
Timeline for enactment of the laws and policies	1970s to 2009 The Treasury Board of Canada has been setting evaluation policy for federal departments and agencies since the 1970s. The 2009 Evaluation Policy was an update of an earlier policy.	1997 to 2004 From 1997 to 2003 the annual budget laws in Chile specified that "social, production promotion, and institutional development programmes included in this budget law may be subject to an evaluation of their results". The law authorised evaluations and specified how these ought to be performed but did not make them mandatory. Regulation of the evaluation system was passed in 2003.	1991 to 2007 A series of laws, decrees and regulations have resulted in a complex legal framework. The Constitution of 1991 mandated the National Planning Department to promote evaluation and performance-based management in the public sector. It also included a requirement for a monitoring and evaluation system. In 1994, the National Planning Department was given legal responsibility for creating the National System for Monitoring and Evaluation (SINERGIA).	2001 to 2011 According to the 2001 Budget Law, all federal programmes subject to operational rules in Mexico require annual external evaluation. In 2004 the Mexican Congress unanimously approved the General Law of Social Development. Its objectives include ensuring the full exercise of social rights, establishing mechanisms to evaluate and monitor social development policy, and to periodically measure poverty.	2001 to 2004 The "No Child Left Behind" Act was enacted in 2001 as an amendment to the Elementary and Secondary Education Act of 1965. The Education Sciences Reform Act was passed the following year, and the Individuals with Disabilities Education Improvement Act in 2004. The No Child Left Behind Act and the Individuals with Disabilities Education Improvement Act were initiated by the Bush administration. The Education Sciences Reform Act was pushed forward by a single Member of Congress.

	Canada	Chile	Colombia	Mexico	USA
Precipitating factors	A process of incremental improvements of policies and recommendations from the auditor general and a standing committee The Auditor General's audit of the former Expenditure Management System called for government expenditure decisions to be anchored by reliable information about the performance of programmes. A subsequent study of the Standing Committee on Public Accounts reinforced the Auditor General's recommendations. The integration of evaluation information into the Expenditure Management System was essential for delivering on budget commitments to establish a system based on principles that included: focusing on results and value for money; consistency with federal responsibilities; and eliminating programmes that no longer serve the purpose for which they were created.	Experience of using evaluation for a national investment system; political responses to public concerns about corruption Several scandals related to financial corruption in 2003 contributed to an agreement between the government and the opposition to support a process of state modernisation, including improvements to the evaluation system. Previous experience was important; Chile had established a National Investment System in 1975 which required pre-investment studies for public investment projects. The logic of evaluating the utilisation of public resources was thus already part of Chilean state culture.	A constitutional emphasis on social accountability, promotion by multilateral agencies and endorsement by the president The construction of the monitoring and evaluation system in Colombia related to the broader historical processes which had resulted in the formulation of the Constitution of 1991 and its focus on social accountability. Multilateral agencies promoted the evaluation of social programmes and President Uribe's administration promoted the enactment of relevant legislation.	Experience from rigorous evaluations of some programmes, a change in government, political pressure from the opposition, and assistance from multilateral agencies Mexico had a number of rigorous evaluations of public programmes already in place. A favourable political context after 2002, following 70 years of single-party control, together with strong political pressure from the opposition contributed to the creation of the National Council for the Evaluation of Social Development Policies (CONEVAL). The Inter-American Development Bank and other multilateral agencies provided technical assistance to Mexico in the early 2000s.	Improvements to previous legislation, initiated by the president and pushed forward by a Member of Congress These laws were not driven by special circumstances but came about as improvements to previous legislation. The Education Sciences Reform Act was pushed forward by a congressman. Professional associations lobbied both for and against the legislation. These laws and amendments were made possible by the people involved in Congress and by the Bush administration's strong focus on education. There was little opposition to the legislation itself.

	Canada	Chile	Colombia	Mexico	USA
Objectives of the laws and policies	**A comprehensive and reliable base of evaluation evidence** The objective of the Canadian Policy on Evaluation was to create a comprehensive and reliable base of evaluation evidence for use in supporting policy and programme improvement, expenditure management, cabinet decision-making and public reporting.	**Increasing the use of evaluations to guide public resource allocations** The increasing emphasis on evaluation in Chile has partly been in response to demands from Congress for more and better evaluations and for the increasing use of such evaluations to guide public resource allocations.	**Multiple objectives** Objectives arising from a complex legal framework including constitutional mandates, laws, and decrees. Broad objectives, such as the improvement of public management, have also been included in documentation from the National Council for Economic and Social Policy (CONPES) documents.	**Inform decisions about public expenditures, as well as measures to make federal programmes more effective and efficient** The monitoring and evaluation system, together with other planning and budgetary reforms, aim to establish results-based financing and management rather than simply procedures for decisions about how to allocate public resources. The purpose of the General Guidelines for the Evaluation of Federal Programmes of the Federal Public Administration is to systematise the monitoring and evaluation of federal programmes in an integrated manner so that the results inform decisions about public expenditures. By doing so, these measures help to make federal programmes more effective and efficient.	**To inform the reauthorisation of programmes** Evaluations are conducted during the period of programme authorisation (typically every 4-5 years) to inform programme reauthorisation by Congress.
Scope of the laws and policies	**All programme spending, including grants and contributions** Large departments and agencies are required to evaluate all their direct programme spending every five years (except where the spending is time-limited), as well as the administration of major statutory spending. Direct programme spending includes operating and capital spending as well as grants and contributions.	**Selected social, production promotion, and institutional development programmes** The sectors and programmes included in the annual evaluations are "the social, production promotion, and institutional development programmes that the Ministry of Finance determines." The scope of the legislation is not comprehensive and is discretionary to some extent. The executive determines how many programmes will be evaluated and how. Only a small proportion of public expenditure has been subjected to evaluation.	**Programmes or policies from every sector that are prioritised by the government** Impact evaluations are not required for every programme or policy implemented. SINERGIA leads around 15 evaluations per year, and roughly 6% of the government spending on programmes that could potentially be evaluated is actually evaluated.	**All government programmes** CONEVAL's scope covers all federal programmes in sectors related to social development. Programmes not related to social development are the responsibility of the Secretariat of Finance and Public Credit. Impact evaluations are required for all programmes, whether they are new or not. However, impact evaluations are not implemented for all programmes. The Annual Evaluation Programme determines which programmes require impact evaluations based on their budget, size and relevance.	**Federal education programmes** For federal education programmes, transparent use of research evidence is used at the launch of a programme and sometimes at the end of a programme.

	Canada	Chile	Colombia	Mexico	USA
Responsibility for implementing the laws and policies	Treasury Board Secretariat The Treasury Board of Canada Secretariat is the central agency responsible for providing leadership for evaluation across the Canadian federal government, giving advice and guidance in the conduct, use and advancement of evaluation practices. Deputy heads of department are responsible for establishing a robust, neutral evaluation function in their department and ensuring that it adheres to the Policy on Evaluation and the supporting directive and standard.	National Budget Bureau The Ministry of Finance must formulate one or more decrees annually specifying which programmes or projects will be evaluated during the year. The Evaluation Programme is part of the Management Control System, located in the National Budget Bureau (DIPRES)at the Ministry of Finance.	National Planning Department In 1994 the National Planning Department was given responsibility for creating SINERGIA and for reporting annually to CONPES (a policy committee headed by the President) on the results of the evaluation system. A National Planning Department resolution in 1994 assigned responsibility for self-evaluation to all agencies in the executive branch of government. The Directorate for Evaluation of Public Policies, a unit established within the National Planning Department, is the technical secretariat of SINERGIA.	CONEVAL CONEVAL has the power – based on the General Law for Social Development – to regulate and coordinate the evaluation of social development policies and programmes and the periodic assessment of whether programmes are complying with their social objectives. The Secretariat of Finance and Public Credit and the Secretariat of Public Service use a performance evaluation system – based on the Federal Budget and Financial Responsibility Law – to identify the efficiency, economy, effectiveness and social impact of public expenditure, and in the case of the Secretariat of Public Service, evaluate the performance and results of institutions. All federal secretariats and agencies are required to adhere to the evaluation guidelines and must use the prescribed monitoring and evaluation instruments.	No overarching body responsible for evaluation There is no overarching body responsible for evaluation. A case-by-case assessment is made for each programme to determine the specific manner in which it is to be evaluated. The Institute of Education Sciences (IES), established in 2002, is the research arm of the Department of Education. The IES is charged with producing rigorous evidence on which to base education practice and policy, with programme evaluation housed primarily in the National Center for Education Evaluation and Regional Assistance.

	Canada	Chile	Colombia	Mexico	USA
Evaluation methods	**No specification of methods** Evaluations methods are not specified. Each evaluation is required to include recommendations, a management response and an action plan.	**Impact evaluations use experimental methods as far as possible** Programme evaluations are based on secondary information sources and logic frames. Impact evaluations are rigorous evaluations which use experimental methods as far as possible. DIPRES manuals define the requirements for each type of evaluation.	**A technical secretariat determines the methods used for each evaluation** Once it has been decided that an evaluation will be undertaken, a committee in the Directorate for Evaluation of Public Policies determines the hypotheses, methods, and relevant technical details. External reviewers are invited to support or discuss the terms of reference for the evaluation.	**Minimum criteria for some types of evaluation and terms of reference are issued by CONEVAL for each impact evaluation** The General Guidelines establish the minimum criteria for some types of evaluation. They do not specify evaluation methods. For impact evaluations, methods are determined through a feasibility analysis and terms of reference. CONEVAL issues the terms of reference for some types of evaluations and these must be observed by federal secretariats and agencies.	**Details of evaluations are negotiated** Sometimes Congress specifies particular evaluation methods and questions that must be answered when authorising a programme evaluation. In other cases they are less prescriptive. Either way, the actual details of an evaluation are a joint product of the descriptions of the work to be accomplished (generated by the Department of Education) and responses from external contractors who compete for the work by proposing how they will translate the Department's general requirements into action.
Funding for evaluations	**Earmarked funds for evaluation** Treasury Board allocations usually receive funds for evaluation and these are typically designated to the evaluation unit of the relevant department. In addition, evaluations are funded through each organisation's budgetary 'reference level' (the general envelope of money approved by parliament that each organisation receives). The amount of funding varies across organisations.	**Evaluations are funded through the DIPRES budget** The Ministry of Finance spends approximately US$0.75 million of the total annual budget of US$20 billion on monitoring and evaluation systems. In 2010, US$2.5 million was allocated to evaluation processes. From 1997 to 2010, 393 evaluations were performed. The proportion of publicly-funded expenditures that are evaluated is low. External funding plays an important role in evaluation funding.	**External funding plays an important role in evaluation funding, but the level of funding from the national government is increasing** Around US$8 million is made available for evaluation annually. In 2010, the cost of evaluations was US$7 million, and approximately US$5 million of this total was provided by international sources.	**A small proportion of programme funds; funds through CONEVAL** In 2007, the costs of 102 of the 106 evaluations coordinated by CONEVAL could be identified. The total cost of US$5.9 million was less than 1% of the budget allocated to the federal programmes evaluated that year. Eight of these evaluations were funded by CONEVAL.	**Earmarked funds for evaluation** There is no independent funding for evaluation in the Department of Education. All funding (US$70 million annually) is provided from money set aside for evaluation in the authorisation of other programmes. The budget allocated for evaluation is small compared to the total US$60 billion budget of the Department of Education.

	Canada	Chile	Colombia	Mexico	USA
Enforcement of recommendations	**Appropriate measures determined by the Treasury Board** Consequences of non-compliance with this policy can include any measure allowed by the Financial Administration Act that the Treasury Board would determine as appropriate in the circumstances. If the Secretary of the Treasury Board determines that a department has not complied with the requirements of this policy or its supporting directive or standard, the Secretary of the Treasury Board may request that the deputy head take corrective actions and report back on the results achieved.	**The most important incentive derives from the fact that the Budget Bureau, which allocates the resources that enabling the functioning of public services, is the evaluation programme manager** Both the previous and current governments have had plans to remove the evaluation function from the budget office and to create an independent agency. However, the Ministry of Finance has rejected this proposal.	**The main mechanism for ensuring that evaluation results are used by decision makers is guaranteeing a good process for an effective evaluation** The legislation is not linked to budgetary, financing or auditing processes. There is no agency responsible for enforcing evaluation recommendations but two offices do have relevant responsibilities. SISCONPES is a system designed by the National Planning Department to monitor responses to recommendations made in CONPES documents. The Directorate for Evaluation of Public Policies monitors compliance in accordance with agreed improvement plans.	**Recommendations derived from evaluations are not binding** There are no mechanisms to identify actions based on evaluation recommendations.	**Evaluation reports do not include recommendations** Legislation does not stipulate how evaluation results should be used by decision makers. The IES is mandated to make evaluation results publicly available. Data must also be made available to external researchers undertaking their own evaluations. Evaluations undertaken by the IES are not intended – and not designed – to include policy recommendations. Deciding how to respond to the results is left up to policymakers and forms part of the wider political process. The IES does not include policy recommendations in evaluation reports because these may undermine the perceived objectivity of evaluations through the introduction of values and political standpoints.

	Canada	Chile	Colombia	Mexico	USA
Evaluation of the laws and policies	Concerns about capacity The Policy on Evaluation is not yet fully implemented. The 2009 Fall Report of the Auditor General reported finding experienced evaluators was a challenge and that extensive use of contractors had been made in order to meet requirements. Departments expressed concern about their capacity to begin evaluating all direct programme spending from 2013, as required by the 2009 Policy on Evaluation. An evaluation of the Policy on Evaluation is planned.	Most recommendations are adopted and implemented The evaluation system maintains detailed information regarding programme changes and the monitoring of recommendations derived from evaluations. Evaluation of the overall system was conducted by the World Bank. An important limitation to the implementation of the requirement for evaluation is the production capacity of the Budget Bureau team responsible for the yearly bidding process. 7% of all the evaluated programmes between 2000 and 2009 were terminated. Over 60% of the recommendations were for minor adjustments or modifications to programme designs or internal processes. Most recommendations are adopted via 'institutional commitments'". Between 1999 and 2007, more than 3,500 institutional commitments were made. Of these, 82% were fulfilled, 11% were partially fulfilled, and 6% were not fulfilled.	Challenges include the absence of clear links between evaluation and planning and budgeting processes The monitoring and evaluation system have public support but the independence and credibility of SINERGIA has been questioned by academic groups and the private sector. Critics argue that as long as SINERGIA is under the control of the executive branch of government, there will be no guarantee of objectivity. Evaluations are nevertheless implemented by external firms. A CONPES policy document about SINERGIA lists four key challenges: the lack of a single, clear conceptual framework; the need to clarify the roles and responsibilities of the organisations supporting SINERGIA; the absence of clear links between the evaluation and the planning and budgeting processes; and problems with the availability and frequency of data and data quality controls.	Evaluation guidelines have resulted in a common language, a learning process, and improved monitoring and evaluation In 2008 the World Bank undertook a study on the implementation of the General Guidelines. It found that the General Guidelines created a common language within the federal public administration and provided a results-based system, that there was a learning process among public servants regarding monitoring and evaluation, and that the provision of a standardised menu of monitoring and evaluation instruments helped to achieve better results. There have been no court challenges to the General Guidelines. Overall, the response to them has been good. Mechanisms should be strengthened so that the results of the evaluations can be linked to decisions about the allocation of resources or to the modification or suspension of policies.	Responsibility for evaluations should be independent of political leadership Legislation has been evaluated and recommendations made. One recommendation is that all evaluations should be the responsibility of the IES. At present, some evaluations are the responsibility of offices that are not independent of the political leadership of the Department of Education.

Enactment of the laws and policies

A range of factors led to the enactment of the laws and policies we identified. In four cases (Chile, Colombia, Mexico, and the USA), they were initiated by new governments as part of a broader set of reforms focused on or motivated by:

- Improving the effectiveness of government policies and programmes
- Improving expenditure decisions
- Improving public management
- A perceived need to improve the system of evaluation in relation to concerns about corruption by public servants
- The need to create body outside of government devoted to evaluating social programmes due to perceptions about a lack of objectivity and technical rigour
- A need for transparency and accountability
- Results-based management

The initiatives were taken, championed or supported by a range of different stakeholders, and included auditors general, budget bureaus, presidents, parliaments, the legislative branches of government, individual Members of Congress, multilaterals and donors. The initiatives built upon earlier laws and policies, existing cultures of evaluation, or past evaluation experience.

In Canada and the USA, these laws or policies were based on the renewal of previous laws or amendments. Experience in other countries did not inform the development of the initiatives devised in the USA. However, Canada researched policies in other countries and drew lessons from other jurisdictions when developing its 2009 evaluation policy. Studies conducted by (or on behalf of) the Treasury Board of Canada Secretariat before the policy renewal, for example, revealed weaknesses in the existing evaluation function that could be addressed through amendments to the Policy on Evaluation.

Similarly, Chile, Colombia and Mexico learned from experiences from other countries and received advice and support from the Inter-American Development Bank, the World Bank and other multilateral agencies. In Mexico, previous attempts had been made to establish results-based management, but very limited evidence was found relating to the implementation process and the extent to which evaluations were used. This necessarily meant that there were limited opportunities to apply previous experience to policy development.

We were able to identify few barriers to the enactment of the laws and policies in the countries that met our criteria. This may possibly reflect the fact that our approach led us to focus only upon those countries in which the enactment and implementation of such laws and policies had already been facilitated. Political opposition or lobbying was not perceived as a major challenge in any of the cases identified, although opposition to some elements of the legislation was noted. Some commentators had expressed unhappiness with the results of specific evaluations but this does

not appear to have affected the overall responses and attitudes towards the requirements for evaluation.

Scope of the laws and policies

Four of the five included cases required evaluations across several sectors. The American case was limited to education programmes alone. None of the laws and policies focused exclusively on impact evaluations and in no instances were impact evaluations required for every programme. The impact evaluations were used across a wide range of sectors. In Colombia, for example, impact evaluations have been recommended for programmes related to social protection, social assistance, information and communication technology, research and development, transport and infrastructure, and the environment.

The advantages and disadvantages of an intersectoral versus a sectoral scope may depend on the size of the country and particular sectors. In the USA, more resources are used annually for evaluation in the education sector alone (US$70 million) compared to the annual expenditure across all sectors in Chile, Colombia and Mexico (which ranged from US$2.5 to US$8 million).

The potential advantages of an intersectoral scope for evaluation include:

- The evaluation methods are similar across the sectors. Centralised intersectoral responsibility can help to ensure quality standards across all sectors
- A specialised, fully dedicated team can be used. Countries with more limited resources may not be able to afford a sufficiently large team within individual sectors, and may therefore benefit from economies of scale
- Independent contractors commissioned to undertake evaluations may be able to move between sectors, such as from education to labour to health care
- Decisions about which programmes to evaluate and how these should be evaluated can be made more independently instead of by those responsible for the programmes. Outside points of view may be beneficial
- An organisation outside specific sectors might help to ensure a more appropriate use of the evaluation results either through monitoring or being engaged in deciding on the recommendations based on evaluation results
- Coverage of all government-funded programmes

The availability of evidence related to policies and programmes from across all public sectors enables governments to make well-informed decisions about how best to use public resources. However, if an organisation responsible for evaluation falls outside the sector responsible for particular programmes, it is important to ensure that it has the necessary communication skills and is able to work well across all sectors. The combination of a broad intersectoral scope and limited resources can result in inappropriate discretion being exercised when deciding which programmes to evaluate and how; many programmes may not be evaluated satisfactorily. From

1997 to 2010, for instance, a small proportion of public expenditure was subjected to evaluation in Chile.

The potential advantages of a sectoral scope include:

- Ownership by those responsible for programmes within the sector and links to cabinet-level responsibility within the sector
- Evaluation results are more likely to be used as a consequence of ownership and linkages
- Increased familiarity with the knowledge base and context of programmes, what people are doing within the sector, and what their needs are
- Starting with sectors committed to evaluation may increase the chances of success and encourage other sectors to build on these successes

A trade-off is therefore evident between the independence and efficiency associated with being outside a sector, versus the greater potential for ownership and communication afforded by being within a sector. However, it is possible to capitalise on the advantages of both approaches. For example, in Canada the Treasury Board's Policy on Evaluation applies to all government spending, but does not preclude other departments from having their own policies. Thus, for instance, Health Canada uses its own policy to make evaluation requirements more specific to the health sector and every department is required to have an evaluation function.

Ideally, all long-term programmes should be evaluated so that informed decisions can be made about whether funding should be continued and whether modifications are needed. In practice, it is not possible to evaluate everything, particularly as financial resources are limited. What there is should not be spread too thinly. Nonetheless, requiring some degree of evaluation or explicit consideration of the need for evaluation for all long-term programmes might help to build a culture in which evaluation is an expected element of government policies and programmes. Knowing that a programme might be subject to evaluation may well push those responsible for a programme to ensure effectiveness.

None of the cases included in this study required the use of systematic reviews of relevant research either related to decisions about whether to launch a new programme or during the design of programme evaluations. Chile focuses on collecting baseline data for new programmes but not on reviews of international evidence. In Colombia, The National System for Monitoring and Evaluation has made efforts to be involved in the design of evaluations in order to make it possible for it to plan for impact evaluations from the very start of public programmes. Attempts have also been made by the System for Monitoring and Evaluation to be involved in the actual design of new programmes so that these are informed by evaluations conducted in Colombia and elsewhere. To this end, they have established a database of evaluations conducted in Colombia: 55 of the 245 evaluations included by the beginning of 2011 had been conducted by the System for Monitoring and Evaluation and 190 by other organisations.

In Mexico, the National Council for the Evaluation of Social Development Policy also has a public database of all the evaluations made under its coordination, extending

from 2007 to 2012. Although systematic reviews are not required to launch a new programme in Mexico, Secretariats and Agencies are required to justify the creation of new programmes, explain the social contribution of these to the strategic objectives of the Secretariat or Agency (to ensure alignment with national planning), and to provide a matrix of indicators.

All of the five cases included in this study were of laws and policies at the national level. In Canada, only programmes at the federal level are included, unless federal-provincial-municipal agreements are established. In Chile sub-national governments execute national programmes with little autonomy. In Colombia, the National System for Monitoring and Evaluation sometimes needs to work with subnational governments when conducting evaluations and greater engagement at this level has been proposed. The National System for Monitoring and Evaluation is exploring how to promote, stimulate and build capacity at the subnational level where the capacity to conduct evaluations is often lacking.

In Mexico, the laws and policies do not apply to programmes at a subnational level of government. However, the National Council for the Evaluation of Social Development Policy has the mandate to provide support to states and municipalities in matters of measuring poverty and evaluation of social programmes and policies, and therefore aims to replicate aspects of evaluation at a subnational level. Some states have enacted social development laws that include requirements for evaluation and established institutions similar to the National Council for the Evaluation of Social Development Policies.

Characteristics of the laws and policies

Who is responsible for implementing the laws and policies?

In the five cases identified, different departments and organisations were responsible for the implementation of the relevant laws and policies. This responsibility is centralised in the Finance Department in Chile, in the Treasury Department in Canada, in the Planning Department in Colombia, and in a decentralised public organisation in Mexico. In the USA, an office within the Department of Education is responsible for implementing laws and policies (see Table 2).

One advantage to locating such responsibilities within a finance department is that this establishes close links to budget decisions. However, in Chile this proximity has not been reflected in formal budgetary impacts. Instead, most evaluations have led to recommendations regarding improvements to programme management instead of recommendations affecting finances. Both the previous and current governments planned to shift this evaluation function from the Finance Department to a new independent agency. Thus far, however, the Finance Department has resisted handing over this responsibility. In addition, the government may find it difficult to accept fully independent evaluations of its programmes, especially new ones to which it is politically committed.

The concept of linking evaluation to budget processes has been adopted in Colombia, and the agency responsible for each programme is now also responsible for linking evaluation results to budgetary decisions. Although the National System for Monitoring and Evaluation has a limited role in the budget process it tries to ensure that recommendations from evaluations are accounted for in the budget.

Ensuring the independence of evaluation is a challenge. At a government level, appropriate strategies to achieve this might include the appointment of an independent office or external board, or the creation of long-term leadership appointments. But a perceived lack of independence may continue to be a problem as long as such responsibility remains within government. Nevertheless, while an organisation outside of government may be seen as more ideal by some, it may not in fact be an effective alternative in relation to how those responsible for public programmes view and use results. Being inside government, for instance, can facilitate communication with those responsible for public programmes, help to ensure that on-the-ground realities are considered in evaluations, and that the results of evaluations are indeed used to inform management decisions.

Both actual *and* perceived independence are important when attempting to assure people that public resources are being used appropriately. But perceived independence may be less important than actual independence when informing programme decisions (for example, on which methods to use and how to report the results of evaluations) and when deciding how to use public funds. Whereas the media and the public might be primarily concerned with issues of accountability, a secondary audience may be more preoccupied with whether informed decisions have been made. There may therefore be good reasons to discuss the results of evaluations with those responsible for public programmes before making these publicly available.

For more established and stable governments, the advantages of evaluations being undertaken within government may outweigh the disadvantages. For less well established or less stable governments, external evaluation may help to ensure both greater actual *and* perceived independence. In both instances, it may also be advantageous to distinguish between an audit function (which might be best placed outside government), and an evaluation function (which might best be based inside government).

How are the laws and policies enforced?

Strategies for enforcing the laws and policies include:

- Having identifiable people or organisations responsible and accountable for:
    - Evaluation
    - Monitoring compliance
    - Taking corrective actions
- Real or perceived power to withdraw funding.

Mechanisms may also be established to ensure that compliance forms part of the general structure of government (i.e. there are clear, understood and accepted re-

sponsibilities and accountability) or via other legislation or policies (e.g. incentives for civil servants that can be used as an incentive for undertaking evaluations).

Challenges to implementing the laws and policies include:

- A lack of skilled people to manage the processes or undertake evaluations
- Inadequate financing
- A lack of routinely collected data or the means to collect reliable data for evaluations
- A lack of awareness of the benefits of evaluation
- The need to build awareness and understanding about implementing a new policy
- Too much discretion being allowed when evaluations are done
- A lack of penalties for organisations that fail to fulfil obligations
- A lack of a clear definition of what a programme is about (i.e. what the focus of an evaluation is)
- Complex legal frameworks caused by multiple pieces of legislation and policies
- Procurement legislation that makes it difficult to commission evaluations

How are evaluations funded?

In all five cases, core funding is allocated to the entities responsible for evaluation. In addition, programme financing includes funds specifically allocated for this purpose, money from external funding, and funds from departments required to pay for evaluations from their own resources.

Canada: Treasury Board fund allocations usually include resources for evaluation and these are typically given to the relevant function within the department responsible. In addition, evaluations are financed through each organisation's general budget.

Chile: Evaluations are funded through the National Budget Bureau. The Ministry of Finance spends approximately US$0.75 million annually on a monitoring and evaluation system. This is a very modest amount compared to its total annual budget of US$20 billion. In 2010, US$2.5 million was allocated to evaluation processes.

Colombia: External funding plays an important role in evaluation funding, but funding from the national government is increasing. Since 2002, the National System for Monitoring and Evaluation spends at least two-thirds of the total annual expenditure allocated to evaluation. This money was provided by the Inter-American Development Bank, the World Bank, and the government itself. In 2010, the cost of evaluations was US$7 million. Of this, US$5 million was provided by international sources.

Mexico: The National Council for the Evaluation of Social Development Policy has an organisational structure of 69 public servants and a budget of almost US$10 million. The Annual Evaluation Programme determines which evaluations will be funded by the Council and which will be funded by Secretariats and Agencies. In

2007, of the 106 evaluations of consistency and results (i.e. evaluation of the consistency of the logframe[2] and monitoring of results) that the Council coordinated, the costs of 102 could be determined. The total cost of all the evaluations was US$5.9 million – a sum which was less than 1% of the budget allocated to the federal programmes evaluated that year. Eight of these evaluations were funded by the Council. Since 2008, the Council provides resources to undertake external evaluations, such as performance-specific evaluations. Other evaluations that Secretariats and Agencies are required to undertake must be funded from their own resources.

USA: No independent funding for evaluation is provided in the Department of Education. All annual funding (US$70 million) comes from funds set aside for evaluation as part of the authorisation of other programmes. The budget allocated for evaluation is small compared to the total US$60 billion budget of the Department of Education.

How are decisions made about which programmes to evaluate?

Approaches to deciding which programmes require evaluation include requiring this for all programmes (but not necessarily impact evaluation) while allowing flexibility in deciding on approaches, or having a structured process for deciding which programmes to evaluate. Structured processes engage different stakeholders using different criteria and processes.

Canada: Large departments and agencies are required to evaluate all their direct programme spending every five years (except where the spending is time-limited). Departments have flexibility in choosing evaluation approaches, designs, methods and data sources, and thus have control over the amount of evaluation resources (i.e. the level of effort) devoted to each evaluation.

Chile: Two processes are used for deciding which programmes require evaluation. In the first, members of parliament may ask for an evaluation during the course of discussions about the budget for the coming year if they are concerned about the performance of particular programmes. These requests are recorded in an agreement protocol. Typically, only a few requests are made in this way. In the second, the National Budget Bureau prioritises which programmes to evaluate by referring to the agreement protocol. 15 to 20 evaluations per year are initiated in this way. No explicit criteria are provided for determining the priorities for evaluation; a number of factors, including the feasibility of the evaluation, can affect the final decisions made.

Colombia: The first step in the creation of the National Evaluation Agenda is the development of a comprehensive list of the programmes and policies that could potentially be evaluated. The prioritisation process that follows includes several steps.

---

[2] The programme logic.

The Directorate for Evaluation of Public Policies first prepares a draft agenda and submits this to directorates in the National Planning Department. These directorates form the link between the Planning Department and the different sectors and are responsible for reviewing, removing, and adding policies. A second draft is then prepared by the Directorate for Evaluation of Public Policies and presented to a committee (consisting of members from the President's Office, the National Development Office and the Finance Ministry) which prioritises the policy evaluations. The amount of resources in the budget of the Directorate for Evaluation of Public Policies is one of the key criteria in determining prioritisation. When insufficient resources are available, a second list of priorities is created and potential funders are consulted. The process ends with a decision being taken about priorities and the National Evaluation Agenda, which is finally approved by the National Council for Economic and Social Policy and the Intersectoral Evaluation Committee.

Prioritisation criteria include budget allocation, the relevance and importance of the programme within the strategic vision of the government, and how long the implementation will take. Topics are then formally included in the National Evaluation Agenda. Following this, the Intersectoral Committee for Evaluation and Management for Results decides what type of evaluation is needed and designs evaluations for the programmes and policies included. Decisions about which activities should be subjected to sophisticated impact evaluations are based on five criteria: the amount of resources they consume, the characteristics of the population they service (e.g. the poor or the displaced), the importance of the activity in terms of whether it is a priority for the national development plan, the innovativeness of the activity, and the potential for replication.

Mexico: The Annual Evaluation Programme is defined jointly by the National Council for the Evaluation of Social Development Policy, the Secretariat of Finance and Public Credit, and the Secretariat of Public Service. This Programme was introduced in 2007 as a planning tool and was formalised in the General Guidelines for the Evaluation of Federal Programs. The Annual Evaluation Programme can be used to mandate evaluations of consistency and results, design, indicators, process, impact, and of specific strategies. The Annual Evaluation Programme determines which programmes require impact evaluations, according to the budget, size and relevance of each programme. According to the Programme, a viability analysis should be conducted by all those undertaking an impact evaluation.

USA: No overarching body is responsible for evaluation. A case-by-case assessment is made for each programme to determine the specific manner in which it is to be evaluated. There is an informal back-and-forth negotiation process in the Department of Education from the time that annual evaluation plans are proposed until they are finalised.

Who does the evaluations?

In the five cases, some evaluations are done in-house but most are commissioned. A key informant noted: "No office in the world has all the expertise that is needed. You always need to hire people with the necessary expertise, particularly with an intersectoral approach."

How are decisions made about what methods to use?

None of the laws or policies identified specifies the methods that should be used and some do not include *any* methodological requirements. Variations are also evident in the extent to which commissioned evaluations specify required methods in the terms of reference. One key informant observed: "The evaluator is a person who has to adapt the methods to the conditions that he finds in the programme being evaluated. I cannot standardise how to use the methods, but I can standardise the steps to go through to conduct an evaluation. Standardised procedures are important."

Canada: The 'Directive on the Evaluation Function' requires all evaluations to address five core issues and include clear, valid conclusions about the relevance and performance of programmes. The five core issues are: continued need, alignment with government priorities, alignment with federal roles and responsibilities, the achievement of expected outcomes, and efficiency and economy. In calibrating the level of evaluation effort, departments take into account their own specific information needs, the programme characteristics and risks, as well as the quality of performance information already available about the programme.

Chile: An International Advisory Panel, comprising professors from the field of impact evaluation, has been established to help improve decision-making regarding which programmes to evaluate and how. The panel proposed the following principles in 2008:

- Extend the use of experimental methods
- Initiate the evaluation process much earlier in project development (piloting)
- Use the best available quasi-experimental methods when experiments are not possible
- Enhance the database of ongoing evaluation through an increase in new data collection
- Increase the links to – and facilitation of – existing administrative data and other forms of data

Colombia: Once a decision has been made to undertake an evaluation, a committee in the Directorate for Evaluation of Public Policies then determines the hypotheses that will be used as well as the methods and relevant technical details. External reviewers are invited to support or discuss the terms of reference for the evaluation.

Mexico: A technical follow-up undertaken by the National Council for the Evaluation of Social Development Policies (CONEVAL), the Secretariat of Finance and Public Credit or the Secretariat of Public Service in Mexico helps to ensure the quality of evaluations. Guidelines establish minimum criteria for some types of evalua-

tion and the methods are specified. For evaluations of consistency and results, design, strategic, performance and processes, the CONEVAL issues the terms of reference that will be observed by the Secretariats and Agencies through a feasibility analysis and the terms of reference. Methods are determined for impact evaluations through a feasibility analysis and the terms of reference.

USA: Legislation defines rigorous evaluation as 'trials', but does not require the use of specific evaluation methods. Methods differ across programmes. The definition of rigorous evaluation as 'trials' has been contested and the redrafting of this legislation has been debated in Congress.

Stakeholder involvement

Those responsible for public programmes were also found to be involved, to some extent, in evaluations in the five cases, although there were variations in how this is achieved. The nature of the involvement of other stakeholders also varies but efforts to involve those outside of government (including the public) are generally limited.

Chile: Evaluation teams no longer include programme officers. This change may have been due to concerns about potential conflicts of interest and instances in which programme officers defended the position of the government. While evaluators continue to interact with programme officers, they no longer form part of the evaluation teams.

Colombia: Wide consultation is required and attempts are made to involve not only the public sector but all relevant stakeholders who have an interest in a programme or policy to ensure greater legitimacy. This is an important way to ensure implementation: stakeholders can make use of lessons learned during evaluations and thereby help to implement changes before evaluations are complete. Decision makers are involved throughout the process of selecting, designing, commissioning, implementing and reporting impact evaluations. Involving stakeholders can require a substantial investment of time and may entail challenges to ensure that there is agreement about the design and quality of evaluations. However, stakeholder involvement can also help to reduce negative attitudes towards the results of evaluations.

USA: An independent advisory board is required for commissioned evaluations to ensure appropriate stakeholder involvement. Concerns have been raised about the role of Congress in this process, owing to its power and the challenge of responding to poorly-formulated questions. Stakeholder involvement is inadequate at the local and state level.

How is the use of the evaluation results ensured?

In all of the included cases, the results of evaluations are made publicly available and disseminated to decision makers. Other ways to improve the likelihood that the results of evaluations are used may include:

- Ensuring the relevance and legitimacy of evaluations
- Designing evaluations to generate not only impact assessment, but hypotheses about ways to improve programmes
- Framing conclusions in a way that will not alienate those responsible for programmes
- Using forums within the legislative and the executive branches and within civil society
- The joint drafting of institutional commitments (both the organisation responsible for evaluation and the organisation responsible for the programme)
- Assigning responsibility to the senior civil servant in each of the relevant departments
- Creating follow-up reports on those aspects of the public programmes that can be improved
- Creating evaluation reports on social development policy which establish recommendations addressed to different decision-makers
- Providing information for budgetary decision-making as part of the performance evaluation system
- Monitoring compliance with commitments

The USA is the only country in which recommendations are *not* included in evaluation reports. Reports may include recommendations for further research, but not for policy. This is because recommendations are seen as introducing values and political standpoints into the evaluations. "Evaluations should be independent and nonpartisan. If recommendations are included (there are always matters of judgement) these will be from a point of view that will antagonise some audiences. There should be no policy recommendations. There are other approaches, e.g. point-counter-point commentaries. Design features of Institute of Education Sciences' reports enhance their use. They are always about the effectiveness of federal programmes. Congress and the Administration are the main audience."

While informants from other countries understood these concerns, they believed that recommendations were still necessary within their settings. One key informant noted:

> *"In a small country it is important to have a hard hand and a soft hand. Evaluations are an opportunity to improve the management of agencies and institutional learning. Recommendations are good for institutional learning. Otherwise, there is not much difference from a tax inspector. In an ideal world you could split these functions. There is not a problem with managers resenting recommendations. There are two stages: first evaluators write recommendations, then [the organisation responsible for evaluation] discusses them with the agency [responsible for the programme]. This is a compromise approach."*

Another informant argued that:

> *"If you don't write a recommendation chapter, the results will be too impersonal, hard to understand. The first person who is going to use the results is*

the political person behind the programme – the public manager. If the evaluation includes a recommendation, the benefits are for him. He needs to understand how the results can be implemented."

But objectivity was still seen as fundamentally important:

"It is important to draw a line between what is and is not an appropriate recommendation. Some recommendations are very objective, based on the results. Those kinds of recommendations are OK. When someone gives recommendations about how to manage the programme he tries not to push too much, it is like telling people how to do their work. There should be clear objective information to support recommendations. Sometimes the evaluators think recommendations should include what they think of the programme. This is not their role. Their recommendations should be based on objective information, not on opinions."

Transparency and independence

Transparency and independence were key themes in all the cases identified. The laws and policies include specific requirements for making evaluation results publicly available or requiring this according to freedom of information legislation. The importance of transparency is also notable in all five cases. However, the real or perceived importance of transparency varies in relation to the different types of decisions, including:

- Which programmes are evaluated
- Who undertakes evaluations
- What methods are used in evaluations
- How the results of evaluations are reported and disseminated
- How evaluations are used (including decisions about recommendations and responses to recommendations, or actions taken in response to the evaluation results)

Ensuring transparency may imply that relevant information should be made open to everyone or, alternatively, that all information is supported by a clear procedure. As one key informant observed:

"Results have to be public, but not in the same moment. The public sector has to understand, analyse and take decisions so they can include those results in their programmes and some results cannot be public in the same moment because of risks for the programme, the public sector and democracy itself."

However, as the same respondent observed: *"There should definitely be transparency all the time."*

In the USA, there is transparency for all the decisions listed above, although one may need to know where to look. Awards for evaluations are publicly available. Evaluation design parameters are publicly available. Reports themselves and proc-

esses are publicly available. Importantly, the data used to generate reports have to be made available to qualified researchers who want access. Any analysis can therefore be examined and reanalysed. "Use of evaluation results is not an issue of transparency. It is just the way things are. It is not something you can see. It is just a chaotic and non-specified process."

The importance of the independence of evaluations is also apparent in all the cases included in this study. However, the degree to which independence (both real and perceived) is seen as critical varies in relation to different types of decisions, including:

- Who pays for evaluations
- Which programmes are evaluated
- Who undertakes evaluations
- The terms of reference for evaluations
- The methods that are used in evaluations
- Reporting and interpreting the results of evaluations
- Peer review of evaluation reports
- How the results of evaluations are used (including who – if anyone – makes recommendations and who decides how to respond to recommendations or the results of the evaluations)

One key informant argued that:

> *"Independence is important, but this does not depend on where you are. For example, independent decisions about methods are important, but do not depend on where you are. Independence in writing the results is also critical. Evaluators should be able to say what they have to say in the best way they can in the results. Those two things are important for independence. Insulation is more important/relevant than independence. An agreement on how you conduct evaluation does not mean a loss of independence."*

However, another commentator from the USA observed:

> *"There is a difference between reports by the Institute of Education Sciences that cannot be stopped and that are not subject to approval by any other federal entity – and [the] "No Child Left Behind" [legislation]. Evaluations that are not [undertaken] by the Institute of Education Sciences are subject to ordinary processes in the Department of Education and reports are routinely put away in filing cabinets. It depends where activities are housed."*

What are the effects of the laws and policies?

The laws and policies have had a variety of desirable effects and no undesirable effects were identified. The evaluations of the laws and policies or systems have lead to improvements in all of the five cases.

Canada: The new Policy on Evaluation aims to provide comprehensive evaluation coverage, better value for money, and more flexibility: programmes can be evaluated

together rather than on a programme-by-programme basis. Under the previous policy, departments traditionally chose to conduct formative evaluations of programmes followed by summative evaluations.[3] Departments are now encouraged to consider which designs could be best customised to suit each evaluation and to meet the information needs regarding the relevance and performance of programmes.

The new policy has not yet been fully implemented and the Treasury Board's Centre of Excellence for Evaluation (responsible for implementing the policy) is obliged to assess the 2009 Policy on Evaluation using the same approach used for programme evaluations. Additional external participants will be involved in the evaluation process to ensure greater neutrality.

Chile: The International Advisory Panel recommended establishing an international advisory committee to review and assess the process periodically. The World Bank has evaluated the overall system but not the legislation *per se*. 7% of all the evaluated programmes between 2000 and 2009 were terminated. Over 60% of all the recommendations made were for minor adjustments or modifications to the design or to internal processes. Most recommendations are adopted via 'institutional commitments' which are closely monitored by the National Budget Bureau. Between 1999 and 2007, more than 3,500 such commitments were made: 82% were fulfilled, 11% were partially fulfilled, and 6% were not fulfilled.

Colombia: The National System for Monitoring and Evaluation is itself undergoing evaluation. This involves assessing the evaluation process, the perceptions of stakeholders, and the effects of the system. The proportion of recommendations that have been implemented based on evaluations is currently being calculated.

Mexico: In 2008, the World Bank reviewed the implementation of the country's evaluation guidelines. The first phase of the study used matrices of indicators to examine the consistency and results of the evaluations. The findings indicated that Mexico's guidelines are focused on:

- The creation of a results-based system offering a common language within the Federal Public Administration
- The establishment of a learning process among public servants regarding monitoring and evaluation
- The provision of a standardised menu of monitoring and evaluation instruments, both for the secretariats and for the programmes, so that better results can be achieved

Several actions were undertaken based on recommendations in the World Bank report. These included:

- Enhancing the timely diffusion of results from public programmes and policy evaluations

[3] The purpose of formative evaluations is to improve programmes, whereas the purpose of summative evaluations is to assess the extent to which a programme achieved intended results.

- Establishing an inter-institutional committee to promote a scheduled interaction and communication between the parties involved with the Guidelines
- Creating an evaluation register

477 external evaluations of various kinds were undertaken in Mexico between 2007 and 2010. 127 of the 158 federal programmes were evaluated during the 2009-2010 term – i.e. just over 80% of those that could potentially be evaluated. No mechanisms were available to identify all the actions undertaken on the basis of the evaluations, but examples include:

- The transformation of the ProTree Programme (a campaign to pant 250 million trees) to incorporate new intervention schemes in order to increase its effectiveness (The changes were based on the results of a 2007 evaluation and recommendations for greater consistency)
- Cancelling the implementation of the Social Security First Job Programme. (The design evaluation found that the programme did not comply with the purposes for which it had been created)
- Recommendations made by the 2008 Evaluation Report on Social Development Policy led to the creation of the Five Steps for Your Health to Live Better programme, an increase to the National Infrastructure Programme, an increase to the Social Security period of coverage, the extension of the Temporary Job Programme to urban areas, and the restructuring of the Procampo operating rules (these establish the way in which a programme will operate, describe the objectives, population, specific support, the actions that will be provided, and the conditions under which people receive the benefit)
- Using information from consistency and results evaluations, performance-specific evaluations, and analysis of programme progressiveness (i.e. distributive incidence of the programmes on income levels) to inform the 2010 budgetary allocation
- Modifying the operational rules of The Rural Housing; Support Fund for Micro, Small and Medium Business; Attention to Adult Education Demand; Clean Water; and Environmental Institutional Development programmes based on findings from performance-specific evaluations
- Using results and management indicators for 491 programmes (including 158 social development programmes) as a result of the continuous coordination between the National Council for the Evaluation of Social Development Policy and secretariats and agencies

USA: Legislation has been evaluated and recommendations made, including making all evaluations the responsibility of the Institute of Education Sciences. Previously, an office in the Department of Education conducted a significant number of evaluations and reported directly to the Secretary of Education. Effectively, this meant that the office did what the Secretary and the administration wanted and resulted in tensions over who exactly was responsible for particular evaluations. The Secretary's office wanted to keep control of those evaluations most important to it, but this meant that the independence of the evaluations could not be guaranteed: those programme officers responsible for implementation were also responsible for the evaluation. The Obama administration took the positive step of clearly demar-

cating responsibilities for evaluation between the Institute of Education Sciences and others responsible for short-term evaluations.

The 'No Child Left Behind' legislation, one of the three laws requiring evaluation, was challenged in Connecticut on the grounds that the funding given to state governments was inadequate and therefore unconstitutional. However, the court decided that the legislation was not a violation and the challenge was finally rejected by the Supreme Court of the United States. This is the only court case we could identify for any of the laws and policies included.

Hundreds of programme evaluations have been conducted in the USA and all large educational programmes are subjected to some form of evaluation. However, a lack of funding to support evaluation, particularly of small programmes, has meant that some programmes which should have been evaluated have not been. Many small programmes have also been discontinued due to budget cuts; evaluation may have provided some protection for small, effective programmes had funds been available to conduct reviews. The Department of Education now pools 2% of funds across small programmes and this is used for evaluation.

Strengths and weaknesses of the laws and policies

The key strengths of the laws and policies include:

- Extensive use of information in budget and decision-making
- Monitoring progress towards political goals
- Active participation of key stakeholders in monitoring and evaluation activities
- Independent evaluation and appropriate levels of financial support
- A strong monitoring and evaluation system
- Improvements in research capacity and quality

Key weaknesses that were identified for one or more of the cases include:

- The need to clarify roles and responsibilities
- A lack of evaluation, oversight and accountability functions
- A lack of comprehensive coverage and discretion in deciding which programmes to evaluate
- Restrictions on how contracts for evaluations are awarded
- Problems with the availability and quality of data
- The absence of clear links between evaluation and planning and budgetary processes
- The low utilisation of the results of evaluations
- Nonbinding recommendations
- A failure to build capacity and disseminate results to subnational authorities
- Requiring evidence-based programmes when there is a lack of such programmes

Arguments against such laws and policies raised by informants include:

- Legal requirements may be meaningless if there is no enforcement
- A legislative approach might be too idealistic and lead to standards that cannot achieve the objective of ensuring that research evidence is used in policy-making in a meaningful way
- Mandatory approaches to encourage evaluation may not necessarily be better than less systematic approaches that are combined with careful performance reviews, or better than indirect approaches
    - Evaluation of public programmes occurs in contexts that may be very different to those settings described in reports. In societies with healthy civil society activities mandating evaluation might not improve upon this
    - The degree of formal evaluation allowed may vary in different political environments
    - Indirect approaches to encourage evaluation may be a better alternative. Programmes that are not evaluated within a particular time, for example, could be re-presented to parliament
- Requiring comprehensive coverage and the evaluation of all programmes may result in unnecessary evaluations (e.g. evaluations of less consequential programmes) at the expense of those that are needed for critical programmes
- Conducting mandatory evaluations on all programmes requires very significant human and financial capital
- Evaluations might become simply bureaucratic requirements that need to be 'checked off'

Discussion

We found only five examples of laws and policies requiring routine evaluation of public programmes. This suggests that the vast majority of countries and international organisations may not be fulfilling their political and ethical obligations to routinely use well-designed evaluations of their policies and programmes when making informed decisions about how to best use available resources to achieve societal goals. Though many nations may in fact be doing so, their laws and policies do not reflect this.

Some countries routinely use systematic reviews for health technology assessments (Appendix 2). However, we were unable to find any examples of laws or policies requiring routine use of systematic reviews to inform decisions about public programmes – i.e. summaries of research evidence that address a clearly formulated question and which use systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies included in the reviews [32]. In some cases the use of available evidence may be a consequence of the way in which such laws and policies have been implemented. For example, 'strategic reviews' are an established feature of the Canadian Expenditure Management System, and in Mexico it is mandatory for Secretariats and Agencies to provide a diagnosis justifying the creation of new programmes. But while these laws and policies require routine evaluation, they do not require the use of systematic reviews or the systematic and transparent use of existing research evidence to inform decisions about the initiation, design, modification or discontinuation of public programmes [32].

A number of important lessons can be drawn from the experience reflected in this review. These include:

- Enacting laws and policies that support routine use of evaluation can:
    - Capitalise on broad reforms (e.g. focusing on accountability and transparency)
    - Build on existing laws and policies (including constitutions, where possible) and evaluation experience
    - Be championed by a wide range people, including auditors general, budget bureaux, presidents, the legislative branches of government, multilateral organisations and donors
- Seeking consensus and political support to promote reforms that foster the use of evaluations, both for the budgetary process and for the improvement of public programmes

- Linking evaluation and monitoring objectives to other complementary government initiatives and institutions to create synergies in budgetary processes, accountability and transparency, in ways that create an integrated rather than overly-regulated system
- Ensuring that reforms to implement a monitoring and evaluation system go hand in hand with administrative reforms that enable those responsible for monitoring and evaluation to respond to the information needs of those making decisions. Monitoring and evaluation should be linked to decision-making
- Focusing clearly on assessing the performance of programmes and core issues which should be addressed in evaluations (e.g. effectiveness, efficiency and equity), and clearly indicating what type of evidence is wanted
- Ensuring that there is an entity with a clear mission to conduct and coordinate evaluation to a high standard, with sufficient insulation from outside pressures that could otherwise impact and influence its work
- Ensuring that trained people carry out evaluations
- Ensuring the gradual implementation of the process
- Building a monitoring and evaluation system using reliable and objective information and continuous improvements to ensure credibility
- Combining mandatory requirements with flexible approaches
    - Discussions should be held about procedures so that people are comfortable about using the correct ones
    - There should be flexibility, for example, in distinguishing between small and large programmes, recognising the nature of the information available, and responding to potential risks
- Recognising that while public policies can be assessed, these assessments may have implications that are not straightforward
- Recognising that an evaluation system has important impacts even if only a small proportion of programmes are evaluated

In all of the included cases, those responsible for the programmes being evaluated were involved in the designing of the evaluations and the interpretation of the results. However, relatively little attention was given to involving stakeholders outside of government who did not have direct responsibility for the programmes. Similarly, several of the included cases reported efforts to use and work with mass media communication and to inform the public about the results of specific evaluations. But none reported efforts to educate the public about the general processes involved, why routine evaluation is needed, or the methods used in rigorous evaluations. Informing the public and engaging a wider spectrum of stakeholders in designing and interpreting the results of evaluations might help to ensure that the evaluation systems created by such laws and policies address questions that are important to the intended beneficiaries and that the results of evaluations are used appropriately. Ultimately this may help to support and strengthen democratic processes.

Three theoretical perspectives are relevant to the interpretation of the findings of this review. The first perspective, Kingdon's three stream model, dominates the literature on political agenda-setting and policy change [33]. According to this model, three separate streams can be found within the arenas of political agenda-setting

and policy change, namely a problem stream, a policy stream, and a political stream. There are moments when "a problem is recognized, a solution is available, the political conditions are right, and the three streams get joined together. So advocates develop their ideas over a long period of time. They develop their rationales and supporting information, they get their proposals ready, and then they strike when such an opportunity comes along." [34] Our findings indicate that there are different types of challenges within the problem stream[4] and that events (e.g. a new government) or broad reforms in the political stream can provide windows of opportunity for advocates of routine evaluation. The experience summarised in this review can help to ensure that appropriate policies are ready, that they are informed by international as well as national experience, and that they can be refined and framed to address the specific problems and political opportunities that arise in specific contexts.

A 'two-communities' or 'two cultures' approach is prominent in the literature on research utilisation. Drawing analogies between the relationship between researchers and policymakers, and the relationship between the natural sciences and the humanities, those who adopt this perspective typically suggest that problems with communication between researchers and policymakers are large and difficult to overcome [35, 36] and that these parties are unable to take one another's realities or perspectives into account. The laws and policies that we have reviewed here illustrate, however, that laws and policies that enable and ensure national evaluation systems have the potential to overcome such challenges [37].

The third key perspective, which also is prominent in the literature on research utilisation, focuses upon the concept of the 'use of research' and attempts to addresses how the word 'use' itself may have fundamentally different meanings. The most frequent categorisations of the different ways in which evidence is used are: direct use ('instrumental' or 'engineering'), selective use ('symbolic' or 'legitimating') and enlightening ('conceptual') [37]. The laws and policies reviewed here are clearly and primarily focused on (and supportive of) the instrumental use of evidence. The extent to which they are able to reduce inappropriate or selective use of evidence may depend in part on the extent to which the independent reporting of evaluation results can be ensured and on the incorporation of adequate strategies to ensure appropriate use of the results of evaluation. None of the laws and policies focused on the enlightening or conceptual use of evidence, but this may rightly be seen as something beyond their scope. One key informant argued that a failure to consider this form of use was a result of the weakness of national legislation. He noted specifically that there are government constraints that prevent funding speculative evaluations of policies and programmes, and that although these evaluations may be risky they are also potentially highly rewarding and valuable. Other such initiatives have fo-

[4] Including a need to improve the effectiveness of programmes, expenditure decisions, public management, the objectivity and rigour of evaluations, or transparency and accountability; or to respond to concerns about corruption.

cused on developing a national research agenda and linking research priorities to policymakers' needs for evidence. It is uncertain what, if any, benefits might be realised by linking such initiatives together with laws and policies requiring routine evaluation of public programmes.

Strengths and weaknesses of this study

To our knowledge, this study is the first to examine the international experience of laws and policies requiring routine evaluation of public programmes. We have attempted to identify contacts in as many countries as possible through personal and snowball networking, and we have complemented our search with a review of literature about the institutionalisation of evaluation (Appendix 1) as well as health, environmental and regulatory impact assessments (Appendix 2). In addition, a draft of this report was widely circulated and reviewed by people from more than 20 countries (Appendix 3). It is possible that additional examples of laws and policies that met our criteria were not included. However, it is unlikely that many examples were missed and we believe that this report reflects the limited experience to date. Other important lessons might be learned by further research examining the experience of countries and international organisations which do *not* have such laws and policies.

Evaluation occurs in many contexts beyond the reach of the laws and policies described in this report: legislative committees hold hearings on policies and laws, executive agencies monitor policy implementation and impacts, courts resolve disputes, and civil society actors evaluate programmes. The five examples that met our inclusion criteria therefore do not represent the majority of evaluations which are not mandated by such laws and policies. Firm conclusions therefore cannot be drawn about the comparative effects of mandated requirements for routine evaluation of public programmes versus non-mandatory approaches.

We relied primarily on written responses from key informants to our questionnaire (Appendix 4) and on a document review. We did not undertake in-depth interviews to inform the design of the questionnaire. Several steps were followed to ensure that the questionnaire captured the information we sought and to check our findings and interpretations. These included preparing summaries of each case, checking with respondents to ensure that our interpretations were accurate, undertaking additional interviews to obtain clarifications and additional information, cross-checking each summary with at least two other informants with different perspectives, asking respondents if any important messages had not been captured by the summaries, and sending a draft of this report to over 50 people for review, including all the respondents and all the members of the advisory group.

Further in-depth interviews might enrich and strengthen the conclusions drawn from this study. Interactive discussions with both the respondents and additional people drawn from a range of relevant backgrounds might also provide further opportunities for reflection upon the experiences summarised in this report and for the

refinement of the lessons that can be drawn from them. More importantly, further experiences should be identified and more rigorous assessments are needed of the effects of laws and policies requiring routine evaluation, including an assessment of their potential undesirable consequences. Although there was little apparent risk of undesirable effects in the five cases identified, a number of arguments have been identified against the need for laws and policies requiring routine evaluation (Box 4).

Box 4. Arguments against laws and policies requiring routine evaluation

General arguments against laws and policies requiring routine evaluation include:

- Evaluation already occurs in many different contexts and is continually used for many issues in societies with healthy civil society activities
- Legislation may be too idealistic: standards may be set that are difficult to achieve. Using research evidence in decision-making and policy-making in a meaningful way may also be challenging
- Indirect approaches may be better, such as 'sunset clauses' [38] requiring the expiration of legislation unless it is reviewed and evaluated
- Inadequate evaluation capacity and a poor evaluation culture may result in the ineffectual implementation of requirements for evaluation and lead to bureaucratic mechanisms that contribute little to transparency or better-informed decisions
- Legal requirements may become meaningless if they are not enforced. Evaluation can simply become bureaucratic requirements that need to be 'checked off' [39]
- Less systematic approaches are not necessarily worse than mandatory approaches

Arguments against requiring evaluation for all programmes – as opposed to requiring the systematic and transparent use of evaluations and requiring assessments of the need for evaluation – include:

- In view of the large number of programmes, conducting periodic mandatory evaluations would require enormous investments of human and financial capital
- Comprehensive coverage may be undesirable if it leads to evaluations being undertaken on less consequential programmes (whether such evaluations are needed or not) at the expense of evaluations of more critical programmes

None of the cases included in this review focused on impact evaluations specifically, although impact evaluations were included in all of them. It is important to recognise that there is a need for different types of monitoring and evaluation: laws and policies help to ensure that appropriate types of evaluation are undertaken to inform different types of decisions. A more detailed focus on impact evaluations might help to ensure that there is better evidence available about the impacts of programmes and policies and, ultimately, that public programmes have desired effects and do not have adverse effects.

What does this study add?

The findings of our review resonate with other review findings about the institutionalisation of evaluation [2-10], particularly related to considerations about linkages to budget processes, the importance of independence, the need for human capacity, and the need to systematise evaluation activities so that they are used routinely rather than on an ad hoc basis. This, to our knowledge, is the first review which has

examined laws and policies requiring routine evaluation and summarised the key lessons that can be learned from these experiences.

Conclusion

Only a small number of countries have enacted and implemented laws and policies requiring routine evaluation of public programmes. Surprisingly, we were unable to identify any international organisations with such a policy.[5] We conclude, based on the limited evidence available, that it would be premature for WHO to develop a framework for formal commitments by governments to improve the use of research evidence using the Framework Convention on Tobacco Control as a model [1]. It would, however, be helpful for WHO or other international organisations to further develop and update the framework outlined in this review based on current experience. This framework could provide the basis for a policy brief [41] which frames the problem appropriately for a specific context, describes options that are informed by international experiences and adapted to the specific context, and identifies barriers to implementing those options together with strategies to address them. Moreover, WHO and other international organisations could serve as practical models themselves by ensuring that they have policies in place requiring them to routinely evaluate their own programmes.

It can be argued that laws and policies that require routine evaluation of public programmes should simply be seen as blueprints, and that what is critical is having people with the necessary skills and motivation to design and carry out evaluations, use the results appropriately, and make the evaluation system work. In the same way that protocols are needed for research, well-designed laws and policies are essential for enabling and ensuring an effective evaluation system. In turn, an effective evaluation system is essential in helping governments fulfil their political and ethical obligations to make well-informed decisions about investments in programmes to improve health, social welfare, education, and justice.

Laws and policies that require routine evaluation should also be routinely evaluated, like other public policies or programmes. Our evaluation in this review indicates that these laws and policies have had important benefits with few or no adverse effects. Future evaluations will be able to strengthen the conclusions that can be drawn from these evaluations by:

[5] The United States Agency for International Development (USAID) published its Evaluation Policy after data collection was completed for this study [40].

- Appraising the types of evaluations undertaken as a consequence of the laws and policies and their appropriateness. Such appraisals should focus particularly on impact evaluations
- Examining the importance (consequences) of recommendations and actions based on the evaluations that are undertaken, again focusing particularly on impact evaluations
- Undertaking cost-benefit analyses of the resulting evaluation systems
- Exploring the potential consequences of specific characteristics of the laws, policies and resulting evaluation systems
- Building on this review and contributing to updated reviews of international experience with requirements for routine evaluation of public programmes, as well as contributing to shared learning

References

1. Oxman AD, Bjørndal A, Becerra-Posada F, Gibson M, Block MAG, Haines A, et al. A framework for mandatory impact evaluation to ensure well informed public policy decisions. Lancet 2010; 375:427-31.
2. Gaarder MM, Briceño B. Institutionalization of Government Evaluation: Balancing Trade Offs. International Initiative for Impact Evaluation, Working Paper 2010; 8.
3. Acevedo GL, Rivera K, Lima L, Hwang H, Eds. Fifth Conference of the Latin America and the Caribbean Monitoring and Evaluation (M&E) Network. Challenges in Monitoring and Evaluation: An Opportunity to Institutionalize M&E Systems. World Bank and Inter-American Development Bank, 2010.
4. May E, Shand D, Mackay K, Rojas F, Saavedra J, Eds. Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean: Proceedings of a World Bank / Inter-American Development Bank Conference. World Bank and Inter-American Development Bank, 2006
5. Mackay K. The Development of Australia's Evaluation System. World Bank, ECD Working Paper Series 1998; 4.
6. Mackay K. Institutionalization of Monitoring and Evaluation Systems to Improve Public Sector Management. World Bank, ECD Working Paper Series 2006; 15.
7. Zaltsman A. Experience with Institutionalizing Monitoring and Evaluation Systems In Five Latin American Countries: Argentina, Chile, Colombia, Costa Rica and Uruguay. World Bank, ECD Working Paper Series 2006; 16.
8. Independent evaluation group: A Diagnosis of Colombia's National M&E System, SINERGIA. World Bank, ECD Working Paper Series 2007; 17.
9. Castro MF, Lopez-Aceved G, Busjeet GB, Ordonez XF. Mexico's M&E System: Scaling Up from the Sectoral to the National Level. World Bank, ECD Working Paper Series 2009; 20.
10. Feinstein O, Zapico-Goni E. Evaluation of Government Performance and Public Policies in Spain. World Bank, ECD Working Paper Series 2010; 22.
11. Harris-Roxas B, Harris E. Differing forms, differing purposes: a typology of health impact assessment. Environmental Impact Assessment Review 2010; 31:396-403.
12. Wismar M, Blau J, Ernst K, Figueras J, Eds. The Effectiveness of Health Impact Assessment: Scope and Limitations of Supporting Decision-Making in Europe. Copenhagen: European Observatory on Health systems and Policies, 2007.

13. Davenport C, Mathers J, Parry J. Use of health impact assessment in incorporating health considerations in decision making. J Epidemiol Community Health 2006; 60:196–201.

14. Wright J, Parry J, Scully E. Institutionalizing policy-level health impact assessment in Europe: is coupling health impact assessment with strategic environmental assessment the next step forward? Bulletin of the World Health Organization 2005; 83:472-7.

15. Krieger N, Northridge M, Gruskin S, Quinn M, Kriebel D, Davey Smith G, et al. Assessing health impact assessment: multidisciplinary and international perspectives. J Epidemiol Community Health 2003; 57:659-62.

16. Mindell J, Joffe M. Health impact assessment in relation to other forms of impact assessment. J Public Health Med 2003; 25:107-13.

17. Phoolcharoen W, Sukkumnoed D, Kessomboon P. Development of health impact assessment in Thailand: recent experiences and challenges. Bulletin of the World Health Organization 2003; 81:465-7.

18. Banken R. Strategies for institutionalizing health impact assessment. ECHP health impact assessment discussion papers. Brussels: European Centre for Health Promotion, 2001.

19. Parry JM, Stevens AJ. Prospective health impact assessment: problems, pitfalls and possible ways forward. BMJ 2001; 323:1177–82.

20. Lock K. Health impact assessment. BMJ 2000; 320:1395–8.

21. Juzwishin DW. Evidence informed decision making in health care: the case for health technology assessment. World Hosp Health Serv 2010; 46:10-2.

22. Pashos CL, Miller DW, Smith MD. Health technology assessment: lessons learned from around the world--an overview. Value Health 2009; 12: (Suppl 2):S1-S5.

23. Banta D, Jonsson E. History of HTA: Introduction. Int J Technol Assess Health Care 2009; 1:1-6.

24. Banta D. The development of health technology assessment. Health Policy 2003; 63:121-32.

25. Wood C. Environmental Impact Assessment: A Comparative Review. 2nd Ed. Essex: Pearson Education, 2003.

26. Wathern P, Ed. Environmental Impact Assessment: Theory and Practice. London: Routledge, 2001.

27. Recent Developments with National and International Environmental Impact Assessment Processes. Proceedings of a Policy Forum, New Orleans, USA, 1997. Ottawa: Minister of Public Works and Government Services, 1998.

28. Radaelli C, de Francesco F. Regulatory Impact Assessment. Chapter 13. In: Baldwin R, Cave M, Lodge M, Eds. The Oxford Handbook of Regulation. Oxford: Oxford University Press, 2010; 279-301.

29. Jacobs S. Current Trends in Regulatory Impact Analysis: The Challenges of Mainstreaming RIA into Policy-making. Washington, DC: Jacobs and Associates, 2006.

30. Rodrigo D. Regulatory Impact Analysis in OECD Countries. Challenges for developing countries. South Asian-Third High Level Investment Roundtable. Dhaka, Bangladesh. 2005.

31. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess RG, Eds. Analysing Qualitative Data. London: Routledge, 1994; 173-94.

32. Oxman AD, Lavis JN, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP). 1: What is evidence-informed policymaking? Health Research Policy and Systems 2009; 7 Suppl 1):S1.

33. Kingdon JW. Agendas, Alternatives, and Public Policies. 2nd edition. New York: Longman, 1995.

34. Kingdon JW. A model of agenda-setting, with applications. Law Review 2001; 2:331-7.

35. Caplan N, Morrison A, Stambaugh RJ. The use of social science knowledge in policy decisions at the national level: a report to respondents. Ann Arbor, MI: The University of Michigan, 1975: 1–63.

36. Snow CP. The two cultures and the scientific revolution. Cambridge: Cambridge University Press, 1959.

37. Innvær S, Vist G, Trommald M, Oxman AD. Health policy-makers perceptions of their use of evidence: a systematic review. J Health Services Res Policy 2002; 7:239-44.

38. Jantz B, Veit S. Sunset legislation and better regulation: Empirical evidence from four countries. Gütersloh: Bertelsmann Stiftung, June 2010.

39. Mackay K. Two generations of performance evaluation and management system in Australia. World Bank, ECD Working Paper Series 11, 2004.

40. USAID. Evaluation: Learning from Experience. USAID Evaluation Policy. Washington DC: USAID, 2011.

41. Lavis JN, Permanand G, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP). 13. Preparing and using policy briefs to support evidence-informed policymaking. Health Res Policy Syst. 2009; 7(Suppl 1):S13.

Appendices

Appendix 1. The institutionalisation of evaluation

The institutionalisation of impact evaluations can be defined as "a process of channeling isolated and spontaneous program evaluation efforts into more formal and systematic approaches" [1]. Experience suggests that institutionalisation is a gradual process which may take years to be successfully completed [2, 3].

Reviews indicate that the institutionalisation of evaluation is influenced by context. Experience from Latin American countries suggests that there are three important elements that may contribute to this: strong stakeholders (for example, a congress or a ministry of finance), evaluation champions, and a democratic system in which the voice of the opposition is heard [1-5]. Access to public information and transparency laws have also been found to be important in facilitating the public disclosure of evaluations [1, 2, 4].

The existence of strong institutional and human capacity, as in Australia's public sector, might help to facilitate the evaluation process and the success of institutionalisation [6]. The building of required capacity and the strengthening of data collection and processing systems are also likely to be important factors in ensuring the success of institutionalisation [1, 2]. Conversely, a lack of adequate expertise and technical capacity may limit institutionalisation efforts.

In Latin American countries, the institutions responsible for evaluation are intersectoral, whereas as in other countries, such as the UK and the USA, they are mainly sectoral. In these latter settings, evaluations are being undertaken within the specific parameters of, for example, the health or education sectors and this may be leading to more evaluation specialisation [2].

The level of independence of the bodies responsible for evaluation varies from case to case. Independence is associated with credibility: the greater the independence, the greater the degree of credibility [1, 2, 3]. CONEVAL's experience in Mexico suggests that having greater independence can also result in greater acceptability outside of government, for example by congress, the mass media, and civil society [2]. Contracting evaluations to external consultants or institutions selected through open bidding processes is one way to ensure greater independence and impartiality [2, 3].

However, if institutions are outside of government and separate from those authorities responsible for budgeting and planning, their ability to plan and implement evaluations and to influence the design and implementation of programmes may be adversely affected [1, 2]. If there are limited links between evaluation institutions and those making programme budget decisions (as is the case in Spain) this too may limit the influence of evaluations on budget processes [7]. Support from central authorities can help to address problems related to the poor enforcement of recommendations [2]. However, the location of evaluation institutions inside or outside government and their links to budget and programme planning are likely to be particularly critical to their success [6]. Similar experience has also been reported by organisations that support evidence-informed health policy, and have pointed to the importance of finding a balance between the need for strong links with policymakers and the need for independence [8].

Systematising evaluation activities – in order to make them a regular and accepted part of political and management life rather than special, ad hoc activities – may also be important. The appropriate timing of evaluations is critical to ensuring that evaluations fit well into policy and decision-making cycles [6]. Involving intended users in the identification of which programmes to evaluate is important to ensuring the relevance of evaluations. It may also increase their sense of ownership and receptiveness towards the findings [3, 6].

The importance of a unique and broad legal mandate has also been pointed out [1].

References

1. Gaarder MM, Briceño B. Institutionalization of Government Evaluation: Balancing Trade Offs. International Initiative for Impact Evaluation, Working Paper 2010; 8.
2. Briceño B, Gaarder MM. Institutionalizing evaluation: Review of International Experience. International Initiative for Impact Evaluation, 2009.
3. Zaltsman A. Experience with Institutionalizing Monitoring and Evaluation Systems In Five Latin American Countries: Argentina, Chile, Colombia, Costa Rica and Uruguay. World Bank, ECD Working Paper Series 2006; 16.
4. May E, Shand D, Mackay K, Rojas F, Saavedra J, Eds. Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean: Proceedings of a World Bank / Inter-American Development Bank Conference. World Bank and Inter-American Development Bank, 2006
5. Feinstein O, Zapico-Goni E. Evaluation of Government Performance and Public Policies in Spain. World Bank, ECD Working Paper Series 2010; 22.
6. Mackay K. The Development of Australia's Evaluation System. World Bank, ECD Working Paper Series 1998; 4.

7. Castro MF, Lopez-Aceved G, Busjeet GB, Ordonez XF. Mexico's M&E System: Scaling Up from the Sectoral to the National Level. World Bank, ECD Working Paper Series 2009; 20.
8. Moynihan R, Oxman AD, Lavis JN, Paulsen E. Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier. Rapport Nr 1-2008. Oslo: Nasjonalt kunnskapssenter for helsetjenesten, 2008.

Appendix 2. Health, environmental and regulatory impact assessment

Different types of impact assessments are routinely conducted in many countries, including health impact assessments (HIAs), environmental impact assessments (EIAs), health technology assessments (HTAs), and regulatory impact assessments (RIAs). An HIA is an assessment of the health effects, both positive and negative, of a project, programme, or policy, and its purpose is to support decision-making [1-3]. An EIA is a procedure through which the environmental impacts of decisions to enact legislation, and to implement policies, plans and projects are assessed [4]. HTAs are used to assess the medical, ethical, social and economic implications of the development, diffusion and use of health technologies [5]. An RIA is a tool with which to examine and measure the likely benefits, costs and effects of new or changed regulations [6].

Reviews of experience related to these different forms of impact assessment can provide helpful lessons for mandatory impact evaluations. We therefore turn now to a brief overview of reviews of such experience and the potential implications for mandatory impact evaluations.

Health impact assessment

The assessment of the health impacts of a programme or policy before it is launched is widely advocated but, until recently, this has not been mandatory. Other than an EU policy requiring a consideration of health, there are currently no legal obligations to conduct health impact assessments. Mandatory health impact assessments have also been established via an amendment to the Thai constitution [7].

Target 14.2 of Health 21 specifies that all countries in the WHO Europe Region are required to have a mechanism for conducting HIAs. The European Commission is also required to check that proposals do not have adverse impacts on health and a checklist approach has been proposed. The Amsterdam Treaty of the European Union states that: “A high level of human health protection shall be ensured in the definition and implementation of all Community policies and activities” [7]. The UK, Sweden, and The Netherlands are among a number of countries that have called for HIAs to be undertaken for policies and major political decisions [8]. From 1993, all Cabinet submissions to the provincial government of British Columbia had to include HIAs. But, after 1996, these guidelines were seen only as optional following a change of government.

Health impact assessments vary greatly in terms of how they are conducted [2]. The more routine use of HTAs may be limited by, among other factors, a lack of agreed methods for measuring health impacts and by gaps in the evidence base for assess-

ing potential health impacts. Some critics argue that there is a gap between the objectives of HIAs and the methods used to conduct them [2].

The following enablers to integrating HIAs into decision-making have been identified [9]:

- Striking a balance between decision-maker ownership and HIA credibility
- Organisational, statutory and policy commitment to HIAs
- The provision of realistic, non-controversial recommendations

A lack of knowledge about the policy-making environment by those undertaking HIAs was found to be the most important barrier to integration. Factors positively influencing the way in which HIAs are conducted included the:

- Use of a consistent methodological approach
- Inclusion of empirical evidence of health impacts
- Timing HIAs such that they are congruent with the decision-making process
- Involvement of expert HIA assessors
- Shaping of recommendations to reflect organisational priorities

Factors affecting the integration process negatively included a lack of standardised methodology, a lack of resources, and the use of jargon.

Legal frameworks may help to establish a permanent awareness of health within decision-making institutions [8]. However, HIAs are seldom undertaken despite the fact that a consideration of health impacts is often a compulsory component in many EIA legal frameworks. The translation of a legal framework into practice might depend on factors such as the existence of an adequate administrative framework [8].

Environmental Impact Assessment

The environmental impact assessment (EIA) process has been used globally for several decades and is often cited as an example of how the assessment process has become successful institutionalised [8]. Prior to the passing of America's National Environmental Policy Act of 1969, there were no specific processes, tools or methods for conducting EIAs. This law was the first to require plausible statements regarding the potential impacts of projects to be disclosed in advance and allowed for the creation of relevant, specific organisations and institutions. This Act is seen as an example of how legal mechanisms can help to rapidly and permanently institutionalise the assessment process [10].

The European Union has established a mix of mandatory and discretionary procedures for the assessment of environmental impacts. The European Union Directive (85/337/EEC) on Environmental Impact Assessments, for example, otherwise known as the EIA Directive, was first introduced in 1985 and amended in both 1997 and 2003. The current Strategic Environmental Assessment Directive (2001/42/EC) allows for the enlargement of the scope of EIAs and the inclusion of the assessment of policies, plans and programmes. The consideration of environmental impacts re-

lating to global and regional impacts has also been included but these are less detailed compared with considerations included in environmental impact assessments of local projects. The EU convention on EIAs in a Transboundary Context (the Espoo EIA Convention), was adopted in 1991 and came into force on the 10th of September 1997 [11]. In Canada, the Canadian Environmental Assessment Act (CEAA) provides the legal basis for the federal environmental assessment (EA) process. Other countries such as Australia, New Zealand, the UK, Columbia, Thailand, and France also have legislation mandating the use of EIAs [12]. Several evaluations of the effectiveness of EIAs have found that they are successful in terms of influencing decision-making [10] but concerns have been voiced that assessments are generally performed when projects have already been decided upon. Such delays mean that EIAs may be unable to influence the choice of options and contribute to achieving the given objectives [13, 14].

Health Technology Assessment

The International Network of Agencies for Health Technology Assessment (INAHTA) defines health technologies as "prevention and rehabilitation, vaccines, pharmaceuticals and devices, medical and surgical procedures, and the systems within which health is protected and maintained" [15]. "An 'HTA' might be a technical evaluation of a medical device done for regulatory purposes; it could be a pharmacoepidemiological study of a drug done or funded by industry for the main purpose of gaining reimbursement; it could be a prospective academic study of the health consequences of a particular health care practice, such as a randomized trial; or it could be a systematic review of any or all aspects of a particular health care practice carried out by an HTA agency" [16].

The use of health technology assessments, or HTAs, has grown remarkably over the last decade, spreading from the USA to Europe and beyond. Formal public HTA programmes associated with ministries of health were first undertaken in Sweden in 1987 and later adopted in Western Europe and in other countries. The driving forces behind this rapid development included: the need to monitor the large expenditures for health care, the visibility of new technologies, and the necessity of rationalising the use of health care technology. During the 1990s, almost all member states of the European Union developed national and regional public HTA agencies and programmes and, in 1993 INAHTA was created (the organisation currently has 46 members from 24 countries). These agencies use generally similar methods (mainly systematic reviews), but the scope of HTAs has also gradually expanded from a focus on the cost-effectiveness of health technologies to an inclusion of broader health system arrangements and social determinants of health [17].

Real or perceived differences in the interests of target audiences may account for between-country variations in the focus of HTAs. HTAs targeted at policymakers, for example, may focus primarily on value for money (i.e. cost-effectiveness), while

HTAs targeted at clinicians may focus primarily on effectiveness and give little or no attention to cost considerations. Researchers, on the other hand, may tend to focus on improving the poor state of research, and industries may tend to focus on issues of efficacy [16].

HTAs in the UK include both systematic reviews and original research (including randomised trials) although the general focus remains upon the presentation of information to inform clinical and management decisions [17]. Founded in 1999, the UK's National Institute for Clinical Excellence (NICE) provides guidance on new technologies and has made a significant contribution to the globalisation of HTAs. Now known as the National Institute for Health and Clinical Excellence, NICE has adopted a transparent review process for determining the impacts of health technologies in relation to their cost to the National Health Service (NHS). Because the NHS is obligated to implement the recommendations of NICE, this confers significant weight to the HTAs it conducts [18].

In Australia, the Pharmaceutical Benefits Advisory Committee evaluates the cost-effectiveness of drugs when advising the Australian Health Minister on whether to place a drug on a list of subsidised drugs, known as the Pharmaceutical Benefits Schedule. In 1993, Australia was the first country to add an economic evaluation to the drug approval regulations which had previously focused only on quality, safety and efficacy [19].

Although HTAs have influenced priority-setting in some health care systems, the impact on priority setting could be described as modest at best [18]. A review of the existing literature on HTA has revealed a lack of evidence on the impacts that HTAs have had on health care budgets, efficiency, and health outcomes [18].

Integrating HTA into the entire spectrum of health care funding and delivery, it has been argued, is vital to increase the chances of HTAs successfully informing policy-making [15]. Banta and Jonsson suggest that in order for HTAs to influence policy and practice, a mechanism is need through which policymakers can be made accountable for their use of evidence [17].

Regulatory Impact Assessment

RIA is a systematic and mandatory appraisal of how proposed legislation will affect stakeholders, economic sectors, and the environment. RIAs are often used to estimate the impact of proposed regulations but they can also be used to examine the effects of current regulations in order to identify effective ways to simplify them or to eliminate burdensome features [20].

RIA has become a global phenomenon since the 1980s due to the extensive pressure for more effective and efficient governance [20]. Approximately 23 of the 30 OECD countries have formal policies mandating the use of RIA in domestic policy-making or have included requirements to conduct RIA within draft legislation. Laws require

the use of RIA in the Czech Republic, Korea and Mexico and they are also required in the United States according to a presidential order, and by prime-ministerial decrees or guidelines in Australia, Austria, France, Italy and the Netherlands. In Canada, Denmark, Finland, Germany, Hungary, Japan, New Zealand, Norway, Poland, Portugal, Sweden and the United Kingdom, the use of RIA has been supported by cabinet directives and decisions, government resolutions, and policy directives. RIA techniques are developing further and RIA is increasingly being adopted by non-OECD countries as well [6].

One of the objectives of RIA is to improve understanding of the impacts of government action, including both the benefits and the costs of action. It can also provide evidence to support "doing nothing" as a real option [6]. Selecting a design model when constructing an RIA programme is difficult and no single form is ideal. The specific context (institutional, social, cultural and legal) of a country therefore needs to be considered carefully when choosing one [6].

To be successful in informing regulatory decisions, RIA must be supported at the highest levels of government, for example by laws or executive decrees [6]. Visibly integrating RIA into the policy process might also be an important way of promoting routine use – for instance by attaching RIAs to legislation sent before parliament. RIAs could also be included in the papers sent to cabinet prior to the consideration of draft legislative proposals [6].

Experience from countries in which RIA has been established for a number of years suggests that RIA might contribute to more efficient regulation and better governance [21].

References

1. Lock K. Health impact assessment. BMJ 2000; 320:1395–8.
2. Kemm J. Perspectives on health impact assessment. Bull World Health Organ 2003; 81:387.
3. Parry JM, Stevens AJ. Prospective health impact assessment: problems, pitfalls and possible ways forward. BMJ 2001; 323:1177–82.
4. Wathern P, Ed. Environmental Impact Assessment: Theory and Practice. London: Routledge, 2001.
5. International Society for Pharmacoeconomics & Outcomes Research (ISPOR). Health Care Cost, Quality, and Outcomes: ISPOR Book of Terms. Lawrenceville, NJ: ISPOR, 2003.
6. Rodrigo D. Regulatory Impact Analysis in OECD Countries. Challenges for developing countries. South Asian-Third High Level Investment Roundtable. Dhaka, Bangladesh. 2005.
7. Mindell J, Joffe M. Health impact assessment in relation to other forms of impact assessment. J Public Health Med 2003; 25:107-13.
8. Banken R. Strategies for institutionalizing health impact assessment. ECHP health impact assessment discussion papers. Brussels: European Centre for Health Promotion, 2001.

9. Davenport C, Mathers J, Parry J. Use of health impact assessment in incorporating health considerations in decision making. J Epidemiol Community Health 2006; 60:196–201.

10. Recent Developments with National and International Environmental Impact Assessment Processes. Proceedings of a Policy Forum, New Orleans, USA, 1997. Ottawa: Minister of Public Works and Government Services, 1998.

11. Convention on Environmental Impact Assessment in a Transboundary Context. Espoo, Finland, 1991. Available online at: http://www.unece.org/env/eia/

12. Wood C. Environmental Impact Assessment: A Comparative Review. 2nd Ed. Essex: Pearson Education, 2003.

13. Bekker M. The politics of healthy policies: redesigning health impact assessment to integrate health in public policy. Delft: Eburon Uitgeverij B.V., 2007.

14. Sukkumnoed D, Kessomboon P, Nusaraporn K, Nuntavorakarn S, Sabrum N. Health impact assessment training manual: a learning tool for healthy communities and society in Thailand, Southeast Asia, and beyond. Bangkok: Research and Development Program on Healthy Public Policy and Health Impact Assessment. Health Systems Research Institute; 2007.

15. Banta D, Jonsson E. History of HTA: Introduction. Int J Technol Assess Health Care 2009; 25 (Suppl 1):1-6.

16. Juzwishin DW. Evidence informed decision making in health care: the case for health technology assessment. World Hosp Health Serv 2010; 46:10-2.

17. Banta D. The development of health technology assessment. Health Policy 2003; 63:121-32.

18. O'Donnell JC, Pham SV, Pashos CL, Miller DW, Smith MD. Health technology assessment: lessons learned from around the world--an overview. Value Health 2009; 12: (Suppl 2):S1-S5.

19. Jackson TJ. Health technology assessment in Australia: challenges ahead '*Medical Journal of Australia* 2007; 187: 262-4.

20. Radaelli C, de Francesco F. Regulatory Impact Assessment. Chapter 13. In: Baldwin R, Cave M, Lodge M, Eds. The Oxford Handbook of Regulation. Oxford: Oxford University Press, 2010; 279-301.

21. Jacobs S. Current Trends in Regulatory Impact Analysis: The Challenges of Mainstreaming RIA into Policy-making. Washington, DC: Jacobs and Associates, 2006.

Appendix 3. Advisory group

Abu Bakar Suleiman, International Medical University, Malaysia

Amir Attaran, Institute of Population Health and Faculties of Law and Medicine, University of Ottawa, Canada

Andrew Podger, Australian National University, Australia

Andy Haines, London School of Hygiene and Tropical Medicine, UK

Arild Bjørndal, Centre for Child and Adolescent Mental Health, Norway

Bertha Briceño, Water and Sanitation Program, World Bank

Bocar Amadou Kouyate, Ministry of Health, Burkina Faso

Carmen Hooker Odom, Milbank Memorial Fund, USA

David Bloom, Department of Global Health and Population, Harvard School of Public Health, USA

David Fidler, Center on American and Global Security, Indiana University Maurer School of Law, USA

Douglas William Bettcher, Tobacco Free Initiative, World Health Organization

Francisco Becerra, Council on Health Research for Development, Mexico

Francisco Songane, Partnership for Maternal, Newborn & Child Health, World Health Organization

Gemma Vestal, Tobacco Free Initiative, World Health Organization

Haichao Lei. Department of Policy and Regulation, Ministry of Health, The People's Republic of China

Haluk Soydan, Hamovitch Center for Science in the Human Services, University of Southern California, School of Social Work, USA

Hassan Mshinda, Commission for Science and Technology, Tanzania

Iain Chalmers, James Lind Library, UK

Judith Healy, Regulatory Institutions Network, Australian National University, Australia

Julia Littell, Graduate School of Social Work and Social Welfare, Bryn Mawr College, USA

Maimunah A. Hamid, Institute for Health Systems Research, Ministry of Health, Malaysia

Manuel Dayrit, Department of Human Resources for Health, World Health Organization

Marie Gaarder, International Initiative for Impact Evaluation, India

Mark Gibson, Center for Evidence-based Policy, USA

Mark Lipsey, Peabody Research Institute, Vanderbilt University, USA

Mark Petticrew, London School of Hygiene and Tropical Medicine, UK

Nelson Sewankambo, College of Health Sciences, Makerere University, Uganda

Octavio Gómez Dantés, Centre for Health Systems Research, National Institute of Public, Mexico

Pierre Ongolo-Zogo, Centre for the Development of Best Practices in Health, Yaoundé Central Hospital, Cameroon

Robert Wells, Australian Primary Health Care Research Institute, ANU College of Medicine, Biology and the Environment, Australia

Ruth Levine, USAID, USA

Sara Bennett, Johns Hopkins Bloomberg School of Public Health, USA

Stella Aguinaga Bialous, Tobacco Policy International, USA

Suwit Wibulpolprasert, Office of Permanent Secretary, Ministry of Public Health, Thailand

Appendix 4. Questionnaire

Mandatory impact evaluation

Should WHO develop a framework for formal commitments
by governments to improve the use of research evidence?

Thank you very much for agreeing to complete this questionnaire.

This is a questionnaire for a survey we are conducting as part of a review we are undertaking on behalf of the WHO Advisory Committee on Health Research (ACHR). The review is intended to inform advice from ACHR as to whether WHO should develop a framework for mandatory impact evaluation.

The objective of the review is to identify jurisdictions or international organisations that have legislation or regulation mandating impact evaluations and to learn from their experience.

This questionnaire is designed to be filled by persons knowledgeable about legislations or regulations mandating impact evaluations; i.e. either obligatory assessment of the likely impacts of public programmes before they are launched or impact evaluations after they are launched or both.

Please do not hesitate to contact us at the following e-mail address if you have any questions regarding this project: Mekdes.Kebede@kunnskapssenteret.no

Please first fill in the information requested in the table below. We will use the address you provide to contact you if we decide to follow up the survey with a telephone interview.

E-mail Address
Telephone
Organisation and current position
Briefly describe how you came to know about this legislation/regulation
Briefly describe, if any, your role in the design or implementation of the legislation/regulation

When, why and how was the legislation or regulation enacted?

1. When was the legislation or regulation enacted?

2. Can you send us a copy or tell us where we can access a copy of the legislation or regulation?

3. Who took the initiative?

4. Who championed the legislation or was instrumental in pushing it forward?

5. To what extent was there organised lobbying activity, and by whom, to shape the legislation and what did this seek to achieve?

6. Why was the legislation or regulation enacted when it was; were there special circumstances?

7. Were examples of other legislation or regulation helpful? If so, which?

8. Was support from international organisations helpful or might it have been helpful? If so, which organisations?

9. What background documents/resources were helpful?

10. What else would have been helpful?

11. Was there other legislation or regulation in place that facilitated development or implementation of the legislation or regulation; e.g. access to public information, transparency or human rights laws?

12. Were there legislative or regulatory barriers to mandatory impact assessments; e.g. that prohibit or limit undertaking evaluations, the design of evaluations, or the utilisation of evaluations?

13. What advice would you give to others considering enacting similar legislation or regulation?

Scope of the legislation or regulation

1. Is systematic and transparent use of research evidence to assess the likely effects of programmes or policies required before they are launched?

2. Are impact evaluations required after they are launched?

3. What sectors are included?

4. What types of programmes or policies are covered?

5. What types of programmes or policies are excluded?

6. What levels of government are included; e.g. does the legislation or regulation cover programmes or policies by sub-national government levels?

7. How are decisions made about which programmes or policies require impact evaluations?

8. Is the legislation or regulation part of a larger reform or policy (e.g. to improve public management, or regulating the provision of public services), which addresses considerations in addition to impact evaluations?

9. What are the main strengths of the scope of the legislation or regulation?

10. What are the main weaknesses of the scope of the legislation or regulation?

Characteristics of the legislation or regulations

1. Is the legislation national, regional, or international?

2. Is the legislation binding and enforceable?

3. What incentives or disincentives does the legislation provide to carry out and use impact evaluations?

4. Does the legislation or regulation include requirements about the methods that are used?

5. What organisation or individuals is responsible for implementing and enforcing the legislation or regulation?

6. To whom are those responsible for implementing and enforcing the legislation accountable?

7. Does the legislation or regulation specify the methods that are to be used in evaluations?

8. To what extent are evaluations done ‘in-house’ (by the organisation responsible for implementing the legislation or regulation) versus commissioned?

9. If evaluations are commissioned, how is this done; e.g. is it done competitively and who is eligible to undertake evaluations?

10. How is the independence of those responsible for impact evaluations ensured?

11. How are those responsible for making decisions involved in designing, implementing and reporting impact evaluations?

12. Is the legislation linked to budgetary, financing or auditing processes; and if so how?

13. What if any mechanisms are there for ensuring that the results of impact evaluations are used by decision makers?

14. Does the legislation or regulation require that the results are publicly available?

15. Does the legislation or regulation require active dissemination of the results or communication with decision makers?

16. How is implementation of the legislation or regulation funded?

17. Roughly how much funding is there annually?

Effects of the legislation or regulation

1. Have the effects of the legislation or regulation been evaluated or is an evaluation planned?

2. Were lessons learnt during implementation and evaluation adopted and used to improve the legislation or regulation?

3. Has the legislation been challenged or used in any court cases?

4. How many evaluations have been undertaken each year since the legislation or evaluation went into effect?

5. What proportion of programmes or policies that could potentially be evaluated has been evaluated?

6. How many changes derived from evaluations has been made?

7. What proportion of recommendations derived from evaluations has been adopted?

8. What proportion of recommendations derived from evaluations has been implemented?

9. What are the main strengths of the legislation or regulation?

10. What are the main weaknesses of the legislation or regulation?

Advocates and critics

1. Who were the strongest advocates of the legislation or regulation?

2. Who are the strongest critics of the legislation or regulation?

3. What has been the response of the mass media to the legislation or regulation?

Examples

1. What are examples of the legislation or regulation being successful?

2. What are examples of the legislation or regulation being unsuccessful?

Other legislation or regulation

Can you suggest other examples of legislation or regulation that we should include in our review (by national or sub-national governments, or international organisations)?

Thank you very much for participating in our survey!

Appendix 5. Case reports for included examples of mandatory impact evaluations

Case Report: The Canadian Policy on Evaluation

Enactment of the policy

On the 11th of April 2006, the Canadian government introduced the *Federal Accountability Act* [1] and The Federal Accountability Action Plan [2] in an attempt to increase its accountability. The passage of this a*ct* led to an amendment to Section 42.1 of the Financial Administration Act. Through the *Federal Accountability Act* and the Federal Accountability Action Plan, the Government of Canada brought forward specific measures that help to strengthen accountability and increase transparency and oversight in government operations. The comprehensive Action Plan includes the *Federal Accountability Act,* and supports policy and other non-legislative measures [2]. The Federal Accountability Act (subsequently through the amendment of the Financial Administration Act), requires every department to review, at least once every five years, the relevance and effectiveness of all ongoing grant and contribution programmes for which it is responsible. However, the law does not stipulate the consequences if a department should fail to fulfil this requirement. As noted by one informant: "What is described are expectations, not reality."

The 2009 Fall Report of the Auditor General of Canada examined how evaluation units in six departments identify and respond to the various needs for effectiveness evaluations [3]. It found that the six departments followed systematic processes to plan their effectiveness evaluations and most of the evaluations that they planned were completed. However, over the audited period (2004 to 2009), each department's evaluations covered a relatively low proportion of its total programme expenses – a range of between five and thirteen percent annually across the six departments. Moreover, many of the effectiveness evaluations did not adequately assess programme effectiveness.

The Canadian Policy on Evaluation was implemented in 2009 [4] and aims to create a comprehensive and reliable base of evaluation evidence for use in supporting policy and programme improvement, expenditure management, cabinet decision-making and public reporting. The Treasury Board of Canada Secretariat is the central agency responsible under the policy for providing leadership for the evaluation function across the Canadian federal government through advice and guidance in the conduct, use and advancement of evaluation practices.

The Treasury Board of Canada has been setting evaluation policy for federal departments and agencies since the 1970s. The 2009 Policy on Evaluation was an update of an earlier policy and formed part of a continuous process of policy improvement. The Policy on Evaluation was renewed in 2009 in support of the implementation of the Expenditure Management System [5]. The Expenditure Management

System, implemented in 2007, is a framework for developing and implementing the government's spending plans and encompasses a number of elements and activities (including planning and evaluation) that guide decisions on the allocation of resources. A key pillar of this system is the ongoing assessment of all direct programme spending, or strategic reviews. One important element of this initiative is the Strategic Review Process, a regular review of the direct programme spending by every department and agency on a four-year cycle. Federal organisations, including departments, agencies and Crown corporations that receive appropriations (public money) from parliament, are required to undertake a strategic review of their direct programme spending and the operating costs of their major statutory programmes on a cyclical basis. Once they have completed a comprehensive review of all of their programmes, organisations are required to identify a total of five percent of their programme spending from their lower performing, lower priority programmes. These funds are then proposed for reallocation to higher priorities.

A clear distinction can be made between the legal requirement for evaluating ongoing programmes of grants and contributions (enacted through the Federal Accountability Act) and the policy requirements established by the 2009 Policy on Evaluation, its supporting Directive on the Evaluation Function, and the Standard on Evaluation for the Government of Canada.

The legal requirement for evaluation pertains specifically to ongoing programmes of grants and contributions, which are types of transfer payments. A transfer payment arrangement is used to transfer monies or make in-kind contributions from the federal government to individuals, organisations or other levels of government (e.g. provincial governments) to further government policy and the objectives of the department.

The policy requirements for evaluation apply to ongoing grants and contributions programmes, and to all other types of ongoing direct programme spending.

Departments are given flexibility in choosing evaluation approaches, designs, methods and data sources, and thus have control over the amount of evaluation resources (i.e. the level of effort) devoted to each evaluation. In calibrating the level of evaluation effort, departments take into account their specific information needs, programme characteristics and risks, as well as the quality of performance information already available about the programme. It is important to note, however, that the Directive on the Evaluation Function requires that all evaluations address five core issues and include clear and valid conclusions about the relevance and performance of programmes. These five core issues are described below:

Core Issues		
Relevance		
1.	Continued need for the programme	*Assessment of the extent to which the programme continues to address a demonstrable need and is responsive to the needs of Canadians*

2.	Alignment with government priorities	*Assessment of the linkages between programme objectives and (i) federal government priorities and (ii) departmental strategic outcomes*
3.	Alignment with federal roles and responsibilities	*Assessment of the role and responsibilities for the federal government in delivering the programme*
Performance (effectiveness, efficiency and economy)		
4.	Achievement of expected outcomes	*Assessment of progress toward expected outcomes (including immediate, intermediate and ultimate outcomes) with reference to performance targets and programme reach, programme design, including the linkage and contribution of outputs to outcomes*
5.	Demonstration of efficiency and economy	*Assessment of resource utilisation in relation to the production of outputs and progress toward expected outcomes*

No specific definition of standards of evidence for evaluations exists, but several agencies mandate Results Based Management and have specific performance-measurement frameworks. The renewal of the government's Expenditure Management System (EMS) in 2007 has led to a greater emphasis on using programme evaluation as an input to expenditure decisions. The renewed EMS, which supports responsible and effective government spending within the fiscal limits reflected in government spending plans, has embedded the use of evaluation information as an important feature. The integration of evaluation information into the EMS was essential for delivering on Budget 2006 commitments to put in place a system based on principles that included:

- Government programmes should focus on results and value for money
- Government programmes must be consistent with federal responsibilities
- Programmes that no longer serve the purpose for which they were created should be eliminated

The renewed system addresses recommendations from the Auditor General's November 2006 audit of the former Expenditure Management System, which called for the government's expenditure decisions to be anchored by reliable information about the performance of programmes. The subsequent study of the Standing Committee on Public Accounts (February 2008) [6] reinforced the Auditor General's recommendations [3] and made its own more precise recommendation that "The Treasury Board Secretariat reinforce the importance of evaluation by adding programme evaluation as a key requirement in the Expenditure Management System."

Evaluation contributes to the three pillars of the Expenditure Management System:

1. **Managing for results**: Evaluations are used by departments on an ongoing basis to "manage to results", i.e. to determine if expected programme results are actually being achieved and thereby to order and inform decisions about continuing, amending, or terminating programme spending.

2. **Up-front discipline**: Evaluation evidence is used in expenditure management to support new spending proposals (such as in the Memorandum to Cabinet process)[6] by helping to compare proposed spending with existing or past programmes and programme results.
3. **Ongoing assessment**: Strategic reviews are comprehensive reviews of all government spending, undertaken over a four-year cycle. Strategic reviews are designed to ensure that all programmes are effective and efficient, are focused on results, provide value for tax money and are aligned with government priorities. These reviews benefit from evaluation, which is a key source of systematic, credible, and neutral evidence on the relevance, effectiveness and efficiency of programmes.

Scope of the policy

Under the Policy on Evaluation, large departments and agencies are required to evaluate all their direct programme spending every five years (except where the spending is time-limited, i.e. where spending is not ongoing), as well as the administrative aspect of major statutory spending (e.g. administrative programme spending under the Canada Pension Plan or the Employment Insurance Program). Direct programme spending includes operating and capital spending. It also includes grants and contributions, which since December 2006 have been subject to a legal requirement for evaluation every five years, as per section 42.1 of the Financial Administration Act. Public debt charges and major transfers to persons and other levels of government are excluded.

Programmes that are not ongoing are excluded. Because the Treasury Board of Canada only has the authority to set management policy for federal departments and agencies, only programmes at the federal level are subject to the policy, unless federal-provincial-municipal agreements are established for the evaluation of programmes that are delivered in collaboration with other levels of government.

Characteristics of the policy

If there is a failure to conduct an evaluation, the policy states that the "consequences of non-compliance with this policy can include any measure allowed by the Financial Administration Act that the Treasury Board would determine as appropriate in the circumstances. If the Secretary of the Treasury Board determines that a department has not complied with the requirements of this policy or its supporting directive or standard, the Secretary of the Treasury Board may request that the deputy head take corrective actions and report back on the results achieved."

---

[6] The transparent use of research evidence was recommended in the preparation of Memoranda to Cabinet. Cabinet discussions precede policy or programme approval.

Each evaluation is to include recommendations, a management response and an action plan. There is no specification of the methods to be applied in evaluations. Whether the evaluations are commissioned or done in-house varies across organisations. If evaluations are commissioned, it is mandatory that the contracting be done on a competitive basis when the amounts are higher than $25,000 (CAD); if the costs are below this threshold, departments may choose to follow a competitive contractor selection method nonetheless. “The person or persons carrying out the evaluation or evaluation-related work must possess or collectively possess the knowledge and competence to carry out the evaluation” [7].

The Policy on Evaluation makes deputy heads of departments responsible for establishing a robust, neutral evaluation function in their department and for ensuring that their department adheres to this policy and to its supporting directive and standard [8]. Deputy heads designate a head of evaluation at an appropriate level as the lead for the evaluation function in the department. They ensure that the head of evaluation has direct and unencumbered access to the deputy head as required. They also ensure that a committee of senior departmental officials (referred to as a Departmental Evaluation Committee) is assigned responsibility for advising the deputy head on all evaluation and evaluation-related activities of the department. The Committee is chaired by the deputy head or a senior level designate and supported by the head of evaluation on evaluation matters. Deputy heads approve evaluation reports, management responses (prepared by programme managers) and action plans, in a timely manner. They are expected to use evaluation findings to inform programme, policy, resource allocation, and reallocation decisions.

They also are responsible for ensuring that complete, approved evaluation reports, along with management responses and action plans, are made easily available to Canadians in a timely manner while ensuring that the sharing of reports respects the Access to Information Act [9], Privacy Act [10], and the Government Security Policy [11]. Furthermore, the Directive on Evaluation (Section 6.1.5.b) requires that heads of evaluation “make approved evaluation reports, along with the management responses and action plans, available to the public with minimal formality and post them on departmental websites in both official languages in a timely manner following their approval by the deputy head” [12].

For any Treasury Board allocation of funds there are usually resources allocated to evaluation. These are typically allocated to the evaluation unit in the responsible department. In addition, evaluations are funded through each organisation’s budgetary ‘reference level’ (the general envelope of money approved by parliament that each organisation receives).

The amount of funding varies across organisations. In the context of grants and contributions, different types of agreements exist. Departments always bear the responsibility for undertaking evaluations of ongoing direct programme spending and generally their evaluation units conduct or commission evaluation work themselves.

Sometimes the non-governmental organisation receiving money for administering a programme is required to collaborate in the conduct of evaluations. Evaluation units have ongoing funding. They are accountable to the deputy minister for each department. When programme funding is approved, it usually includes funding for evaluations. This money is usually transferred to the control of the head of evaluation.

The provision of resources to support the implementation of the evaluation requirements is regarded as very useful by the agencies, outside organisations and departments which are required to undertake evaluations.

Effects of the policy

The new policy is not yet fully implemented. The departments examined in the 2009 Fall Report of the Auditor General [3] reported that it was a challenge to find experienced evaluators and that they had made extensive use of contractors to meet their requirements. Departments expressed concern about their capacity to evaluate all direct programme spending from 2013, as required by the 2009 Policy on Evaluation. To ensure full coverage (which includes grants and contributions), they will have to evaluate an average of 20% of their direct programme spending in each year of the five-year cycle.

An evaluation of the Policy on Evaluation is planned by the Treasury Board's Centre of Excellence for Evaluation, which is responsible for leading, supporting and monitoring departments as they implement this policy. An evaluation of the policy is required and the same approach will be taken as is generally required for programme evaluations. Others participants will be involved in the evaluation process to ensure neutrality.

References

1. Canadian Minister of Justice. Federal accountability act. 2006. http://laws-lois.justice.gc.ca/PDF/Statute/F/F-5.5.pdf
2. Treasury Board of Canada. Federal accountability Action Plan. 2006. http://www.tbs-sct.gc.ca/faa-lfi/docs/ap-pa/ap-patb-eng.asp
3. Auditor General of Canada. 2009 Fall Report of the Auditor General of Canada. 2009. http://www.oag-bvg.gc.ca/internet/English/parl_oag_200911_01_e_33202.html#hd3a
4. Treasury Board of Canada. Policy on Evaluation. 2009. http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15024
5. Treasury Board of Canada. The strategic review process. http://www.tbs-sct.gc.ca/sr-es/index-eng.asp
6. Public Accounts Committee. The Expenditure Management System at the Government Centre and the Expenditure Management System in Departments. Retrieved December 21 2010 from

http://www2.parl.gc.ca/HousePublications/Publication.aspx?DocId=3291523&Language=E&Mode=1&Parl=39&Ses=2&File=12

7. Treasury Board of Canada. Standard on Evaluation for the Government of Canada. April 1, 2009; 6.3.2 a. http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15688
8. Treasury Board of Canada. Annex A - Core Issues to be Addressed in Evaluations. In Directive on the Evaluation Function. http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15681&section=text#appA
9. Department of Justice, Canada. Access to Information Act. http://laws.justice.gc.ca/en/A-1/
10. Department of Justice, Canada. Privacy Act. http://laws.justice.gc.ca/en/P-21/
11. Treasury Board of Canada. Government Security Policy. http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=12322
12. Treasury Board of Canada. Directive on the Evaluation Function. http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15681

Enactment of the legislation

Between 1997 and 2003, regulations in Chile specified that "social, production promotion, and institutional development programmes included in … budget law may be subject to an evaluation of their results". This legal authorisation specified how the evaluations of programmes ought to be performed, but did not make such evaluations mandatory.

An amendment of the Constitutional Law of General Principles for Administration of the State issued in 2003, however, a modification of the original law,[7] which declared that "the social, production promotion, and institutional development programmes that the Ministry of Finance determines by one or more decrees *must* be subject to annual evaluations". Effectively, this meant that *all* government-funded programmes became subject to evaluation and that the findings and conclusions of these evaluations had to be reported both to the Chamber of Deputies and the Senate and, after 2004, to the 'Comisión Mixta', a special parliamentary commission established to consider budget issues.

The regulations of the Programme Evaluation System established by Law 1,263 were enacted by the Ministry of Finance's Decree 1,177 of 2003. These required the Ministry of Finance to formulate one or more decrees annually, specifying which programmes or projects will be evaluated during the next year. Parallel to these discussions, negotiations between the executive branch of government and the parliament led to the issuing of Agreement Protocols that defined which programmes and projects would be accepted for evaluation the following year.

From 2004, a number of decrees (especially Decree 1,117) were used to identify programmes or projects for evaluation, and these were included in an official document issued by DIPRES (the budget department of the Ministry of Finance) entitled 'Instructions for the Execution of the Public Sector Budget Law'.

Implemented in 1997 by the Lagos administration, these initiatives were not the result of lobbying, but of strong support both from the Budget Bureau of the Ministry of Finance and the government. This, despite the fact that circumstances at the time were not conducive to the initiation and regulation of evaluation processes. In 2003, faced with scandals over fiscal resource mis-utilisation, the government reached an agreement with the opposition to support a process of state modernisation. Reforms included the creation of a system of high-level public administration and the implementation of a new law for the financing of political parties. The regulation of the evaluation system in Chile may therefore be seen as the product of public concerns about corruption by public servants and the perceived need to improve the state's systems of evaluation.

[7] Decree 1,263 of 1975, which modified section 52 of the original law by means of Law 19, 896.

The Evaluation of New Programs, created in 2008 is a new evaluation system which focuses on two key objectives:

- The design of an evaluation at the beginning of each new programme
- The establishment of control groups for the evaluation, based on randomised trials whenever possible

This system, in turn, is supported by the International Advisory Panel, [8] consisting of experts from the field of impact evaluation field and is responsible for periodically reviewing and assessing evaluation processes for newly selected programmes. The panel observed: "The increasing emphasis on evaluation within the Chilean context has been in part in response to demands from Congress for more and better evaluations and for the increasing use of such evaluations to guide public resource allocations" [1].

International examples, advice and experience also informed the development of Chile's evaluation system. The Budget Bureau team, which is responsible for implementing the DIPRES legislation, paid particularly close attention to previous international programme examples such as the British 'Next Step' and 'Value-for-Money' systems, as well New Zealand policies. Collaboration and contact with institutions such as the Organisation for Economic Cooperation and Development (OECD), and seminars with experts from the OECD and World Bank were also important.

Previous evaluation experience from within Chile was also influential: the logic of evaluating the utilisation of public resources has been a part of Chilean state culture for considerable time. Chile, for example, pioneered the use in Latin America of a National Investment System; pre-investment studies as requirements for public investment projects were also implemented as long ago as 1975 and have since been 'exported' to several other countries in the region. The Legislative Decree 1,263 of 1975 states that the Constitutional Law of General Principles for the Financial Administration of the State requires that "pre-investment studies and investment projects as internal document of the Administration must have a technical-economic evaluation by a national or regional organization (whichever corresponds), based on a technical-economic evaluation that analyzes the project's profitability. The Ministry of Finance shall impart the corresponding instructions and resolutions."

In 1997, the evaluation system was extended to include public programmes and institutions and three instruments were applied: governmental programme evaluations (1997); impact evaluations (2001); agency evaluation (2002) and evaluation of new programmes, introduced in 2009 and aimed at expanding impact evaluations [1].

---

[8] Professors Jere Behrman (University of Pennsylvania), Orazio Attanasio (University College of London), Paul Gertler (University of California, Berkeley), Petra Todd (University of Pennsylvania). It includes local participants as well: professors David Bravo and Claudia Martinez (University of Chile).

The Transparency Law 20,285 of 2008, passed five years after the initiation of Chile's impact evaluation system focuses upon the importance of:

- The transparency of the public service
- The right to access information from public administration bodies
- Procedures for exercising this right and protecting it
- Exceptions to the disclosure of information

This law helps to ensure that members of the public are given access to the results of evaluations, and that all evaluation information generated by DIPRES is publicly disclosed [1].

However, the implementation of such requirements is limited by the production capacity of the Budget Bureau team responsible for the yearly bidding process. It is important, therefore, that any government or organisation that is considering similar legislation or regulations should ensure the gradual implementation of evaluation processes while, at the same time, increasing the level of independence of the evaluations themselves from governmental institutions.

Scope of the legislation

The focus of the current system in Chile is on the appropriate design of evaluations. This is because previous evaluation systems suffered from a lack of baseline information and robust design [1].

"The social, production promotion, and institutional development programmes that the Ministry of Finance determines by one or more decrees" are evaluated annually at a national level. In Chile sub-national governments execute national programmes with little autonomy.

Two processes are used in Chile for deciding which programmes require evaluation. In the first, the Budget Bureau prioritises programmes by referring to agreement protocols. No explicit criteria are provided for determining the evaluation priorities and a number of factors, including the feasibility of the evaluation, can affect the final decisions made. Between 15 and 20 evaluations are initiated in this way each year. In the second process, members of parliament, during the course of discussions about the budget for the coming year, may ask for an evaluation to be undertaken if they are concerned about particular programme performances, and record their recommendations in an agreement protocol. Typically, only a few requests are issued in this way each year.

Chile's current evaluation legislation is not comprehensive and some regard it as too discretionary. At present, the executive division of government determines how many programmes will be evaluated and how this will be done. The International Advisory Panel, noted earlier, was established to support the evaluation of new programmes and proposed the following guidelines in 2008 [1, 2]:

- Extend the use of experimental methods
- Initiate the evaluation process much earlier in project development (i.e. through the use of piloting)
- Use the best available quasi-experimental methods when experiments not possible
- Enhance the database of ongoing evaluation by increasing new data collection and facilitate the use of existing administrative as well as links to other data
- Ensure transparency and integration with budgetary processes
- Establish an international advisory committee to periodically review and assess the process

Characteristics of the legislation

The legislation is binding and enforceable. The fact that the Budget Bureau is the evaluation programme manager *and* resource allocator (enabling the functioning of public services) is a powerful compliance incentive. Full cooperation with DIPRES process is therefore largely assured.

Decree 1,177 issued by the Ministry of Finance stipulates that government programmes should be subject to both programme evaluation and impact evaluation. Programme evaluations are based on secondary information sources and logic frameworks for programmes. For these evaluations the evaluation team reviews the logic framework for each programme and available indicators. Impact evaluations, on the other hand, are rigorous evaluations using experimental methods so far as possible. Both types of evaluation are defined in more detail in DIPRES manuals at http://www.dipres.cl/572/propertyvalue-15409.html. Every evaluation that is tendered must comply with the terms of reference specified.

DIPRES forms part of the management control division of the Budget Directorate. The head of the management control division reports directly to the Budget Director who, in turn, reports to the Minister of Finance. The Budget Directorate is accountable to members of the Chilean Congress who, via the agreement protocols noted earlier, have a say in the approval of the programmes selected [1].

Evaluations of government programmes are performed by independent consultants who are typically either academics or private consultants. Impact evaluations are also performed by consultants or by universities after being put out to competitive tender. Requirements for all evaluations are specified in the terms of reference and a score is assigned to each application based on the specified selection requirements.

Evaluations are not performed by the institution responsible for the execution of the programme being evaluated. Instead, the independent evaluations are performed by panels of experts or by consulting firms from outside the public sector or universi-

ties, and these companies are hired and managed by the Ministry of Finance [3]. Beginning in 1997, programme officers were included in evaluation teams. However, this practice was discontinued due to concerns about potential conflicts of interests and programme officers sometimes defending the position of the government. While evaluators continue to interact with programme officers, they are no longer included in evaluation teams.

DIPRES reviews forms, report procedures, and other standard products for the evaluations of government programmes. Its approval is a prerequisite for payment.

DIPRES has claimed that evaluations are linked to budget processes, given that the evaluations form part of the commitments outlined in the Legislative Power-Executive Agreement Protocol and the Budget Law. However, in reality such links with the budget are weak. Most recommendations relate to management improvements and to changes to the programmes and institutions: these do not often directly affect their budgets or even relate to them.

At the conclusion of an evaluation, the Budget Bureau together with the institution or programme that has undertaken the evaluation issue a document containing 'Institutional Commitments'; i.e. the improvements that the evaluators believe an institution or programme should be required to introduce. This Commitment Document establishes the terms and indicators that will be monitored by DIPRES to ensure that the results of evaluations are used. The only enforcement mechanism available to DIPRES is to threaten potential budget reductions, although in practice such threats are not followed through. Nonetheless, most commitments are achieved. This may be due, in part, to the influence of the Budget Bureau due to its power over government programmes.

The evaluations are reported to the Budget Directorate, to Congress and the public, and they are made available via the DIPRES website (http://www.dipres.cl/572/propertyvalue-15697.html). All completed evaluations, including the full reports and executive summaries, are also publicly available.

Evaluations are funded through the DIPRES budget [1, 4]. The Ministry of Finance spends a total of approximately US$0.75 million annually on monitoring and evaluation – a very modest sum compared to its total annual budget of US$20 billion [5]. In 2010, $2.5 million was allocated to the evaluation process. From 1997 through to 2010, 393 evaluations were performed and only a small proportion of public expenditure has been subjected to evaluation.

Effects of the legislation

The evaluation system provides detailed information regarding programme changes and monitoring of recommendations derived from evaluations [1], but an evaluation of the overall system, such as the one undertaken by the World Bank, is not legally prescribed.

Programme evaluations seldom provide firm conclusions, and those that are given may vary according to the quality of the data generated. Problems may also be described in diplomatic rather than direct ways and mention may be made, for example, of 'challenges' rather than 'problems'. Although this indirect approach may obscure the exact nature of the problem described in evaluation reports, it may also help to reduce antagonism towards the evaluation system and thus facilitate the realisation of the changes needed. Institutional commitments which are made tend to address problems even when these have not been explicitly spelled out in the evaluation reports.

Only 7% of all evaluated programmes between 2000 and 2009 were terminated and more than 60% of the recommendations made have been for minor adjustments or modifications to programme designs or internal processes. Most recommendations are adopted via the Institutional Commitments, which are closely monitored by DIPRES [1]. Between 1999 and 2007, 82% of the more than 3,500 Institutional Commitments were fulfilled, 11% were partially fulfilled, and 6% were not fulfilled [1].

Advocates and critics of the law

The existing legal framework is strongly advocated by DIPRES and by many within the academic and political spheres. Although no major criticisms of the existing system or legislation have been identified, specific evaluations, methodologies and procedures applied in specific evaluations have been perceived negatively.

Strengths and weaknesses of the legislation

One of the main strengths of the monitoring and evaluation system in Chile is the fact that the information obtained is used extensively in budget analysis and decision-making. It is also used to impose programme improvements as well as in the reporting of government performance both to the Chilean Congress and to the public. However, according to the World Bank review of 2005 [6], the utilisation of this system by ministries and agencies (beyond those on which it is imposed directly by the Ministry of Finance) [5] is limited. This, it has been argued, is due to a poor sense of ownership and a lack of acceptance of the evaluation system itself – products of the centrally-driven nature of the wider government system and because of the absence of incentives to conduct evaluations [5].

Both the previous and the current government have had plans to remove the evaluation function from the budget office and to create an independent agency. However, the Ministry of Finance has resisted handing this evaluation function to an independent agency. It may also prove difficult for the government to accept a fully independent evaluation of government programmes, especially of new programmes to which the government is politically committed.

References

1. Bertha B. Institutionalizing evaluation: Revisiting three leading models in Latin America. 2009
2. Briceño B, Gaarder MM. Institutionalizing evaluation: Review of International Experience. Egypt: International Initiative for Impact Evaluation, 2009.
3. May E, Shand D, Mackay K, Rojas F, Ssvedre J (eds). Towards the Institutionalization of Monitoring and Evaluation Systems in Latin America and the Caribbean. World Bank/IADB, 2006. http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/LACEXT/0,,contentMDK:20893139~pagePK:146736~piPK:146830~theSitePK:258554,00.html
4. Independent evaluation group. Experience with Institutionalizing Monitoring and Evaluation Systems In Five Latin American Countries: Argentina, Chile, Colombia, Costa Rica and Uruguay. ECD working paper 16. World Bank, 2006.
5. Macka K. How to Build M&E systems to support better government. World Bank. 2007
6. Rojas F, Mackay K, Matsuda Y, Shepherd G, Azul del Villar, Zaltsman A, Krause P. Chile: Study of Evaluation Program—Impact Evaluation and Evaluations of Government Programs. Washington, DC: World Bank, 2005.

Enactment of the legislation

The development of the monitoring and evaluation system in Colombia is closely associated with broader historical and political changes and, in particular, to the establishment of the national constitution in 1991. Emphasising the participatory nature of Colombia's democracy and the new role of social accountability [1, 2], the constitution has provided a mandate for the National Planning Department to implement evaluation and performance-based management in the public sector [1], and the establishment of enforced monitoring and evaluation (article 343 and 344).

In 1994, the National Planning Department was given legal responsibility for reporting evaluation results annually to the National Council for Economic and Social Policy (CONPES) headed by Colombia's President. SINERGIA, the national system for monitoring and evaluation was created in 1994 and served to operationalise the constitutional and legal mandates for evaluation by assigning responsibility for self-evaluation to all agencies in the executive branch of government [2]. The Directorate for Evaluation of Public Policies, a unit established within the National Planning Department, became the technical secretariat of SINERGIA [1]. According to a law passed in 2003 (Ley 819 de 2003 o Ley de Responsabilidad Fiscal), the national budget is now required to include details of the objectives, intended results, and management indicators for all government activities [2].

In addition, a series of laws, decrees and regulations were passed by the Colombian government to further delineate elements of the evaluation system (e.g. decree 2167 in 1992, CONPES 2688 in 1994, resolution 063 in 1994, CONPES 2790 in 1995, CONPES 2917 in 1997, CONPES 3100 in 1999, decree 1363 in 2000, CONPES 3106 in 2001, CONPES 3117 in 2001, CONPES 3294 in 2004, Art 132 in Law 1151, 2007and CONPES 3515 in 2008) [1]. Much of the legislation that regulates SINERGIA does not form part of larger reform or policy. The inclusion of recommendations for evaluation within CONPES documents, for example, may be the result of national norms established within a range of legislation, including constitutional mandates, laws, and decrees. This legislation may have broad objectives, such as improving public management. The result is a complex legal framework for evaluation in Colombia [2].

Factors that were important in the development of Colombia's current evaluation system include:

- The promotion of the evaluation of social programmes by multilaterals. Since the late 1990s, donors have increasingly demanded evidence regarding the efficacy of development projects [1].
- The endorsement by President Uribe's administration of the evaluation system and the information it provides. This allowed for a resurgence of interest in evaluation following a period of stagnation during the late 1990s

[1]. The focus of this renewed interest was initially on social programmes, such as the evaluation of conditional cash transfers.

Scope of the legislation

Impact evaluations are not required for every implemented programme or policy. However, some CONPES documents – which outline the design of programmes or policies of importance for government priorities – also require the relevant Ministries or agencies to undertake impact evaluations. The development of the National Evaluation Agenda aims to comprehensively review policies and programmes and to assess factors such as the feasibility, relevance and the potential results, impacts and adequacy of an evaluation. SINERGIA aims to include programmes or policies in the National Evaluation Agenda from every sector that has been prioritised by the government's National Development Plan. Some CONPES documents now also recommend impact evaluations in many sectors, including social protection, social assistance, information and communication technology, research and development, transport and infrastructure, and the environment.

Inclusion criteria (listed in CONPES documents or other evaluation agendas) are significantly more important than *exclusion* criteria. CONPES documents may indicate that the prioritisation of an evaluation agenda must be given approval by a specialised (government) committee but such committees may also have their own criteria. Donor-funded programmes may also be prioritised, for instance, because of international agreements. If a public policy is included in the National Development Plan (submitted to Congress every four years at the beginning of a presidential term) then it will also be prioritised for evaluation. The amount of resources required by a policy or programme may also be used as a criterion, although such considerations are regarded as less important than the impacts of the programme on national development.

Hence, at first glance no programme or policy type is excluded from being eligible for a CONPES document or from being included in the Evaluation Agenda.

CONPES documents are policy documents issued by the government and therefore generally cover government programmes or policies at a national level. CONPES documents, however, are developed via discussion and consultation with many actors and sometimes recommendations in CONPES documents include recommendations for engagement in evaluation initiatives directed at the subnational level of government.

The Intersectoral Evaluation Committee (created by decree in December 2002) has responsibility for the coordination, monitoring and evaluation of activities among government units, and deciding how the four-year evaluation agenda and corresponding annual agenda of evaluations should be conducted [2]). CONPES documents, in turn, specify the programmes or policies that require an evaluation and

should be included in the National Evaluation Agenda. Other forms of evaluation, such as process evaluations and methods recommendations, may also be included.

The first step in the creation of a National Evaluation Agenda is the development of a comprehensive list of evaluable programmes and policies. This is done by creating a database based on reviews of CONPES documents, external credit operations, a database of investment projects, and other policy documents. This portfolio is essentially a public investment management system and contains a collection of all capital projects conducted during a fiscal year at a national level. It is also used to identify more resource-demanding projects.

A multi-step prioritisation process is then followed. The Directorate for Evaluation of Public Policies prepares a draft agenda based on the above-mentioned criteria and submits this to different directorates in the National Planning Department. These directorates act as links between the Planning Department and different sectors which will review, remove, and add further policies for evaluation. After preparing a second draft, this agenda is presented to a committee which includes members from the President's Office, the National Planning Department, and the Finance Ministry and a decision is taken about which policies to prioritise and evaluate. At this second stage of assessment, the amount of resources in the budget of the Directorate for Evaluation of Public Policies is one of the key criteria for prioritisation. If insufficient resources are available, a second tier of priorities is defined and potential funders are consulted.

The prioritisation decisions taken are then approved by CONPES and the Intersectoral Evaluation Committee using the budget allocation, the relevance and importance within the strategic vision of the government, and the time of implementation as key prioritisation criteria. After being formally included in the National Evaluation Agenda, the topics are then assessed by the Intersectoral Committee for Evaluation and Management for Results who decide what type of evaluation is needed and which design evaluations to use for the included programmes and policies. Decisions about which activities should be subjected to sophisticated evaluation are based on five criteria, namely the: "(a) amount of resources they consume; (b) the characteristics of the population they serve (e.g. the poor, or the displaced); (c) importance of the activity, in terms of whether it is a priority for the national development plan; (d) innovativeness of the activity (e.g. a pilot); (e) potential for replication" [2].

Once it has been agreed that an evaluation will be performed, a committee in the Directorate for Evaluation of Public Policies determines the relevant hypotheses, methods, and relevant technical details. External reviewers are invited to support or discuss the terms of reference.

The wide consultation required in this process of strategic evaluation is a key strength of the system: all relevant stakeholders with an interest in a programme or policy are consulted and attempts are made to include perspectives from both within the wider civil society and the public sector. This is an important way to ensure

greater legitimacy and offers stakeholders the opportunity to check whether the consultancy firm conducting the evaluation is following the correct procedures and considering their concerns. It is also an important way to ensure implementation, because stakeholders can make use of the lessons learned during the evaluation and help to implement changes before the evaluation is complete. Stakeholder involvement can also help reduce negative attitudes towards the results of an evaluation.

Characteristics of the legislation

No single agency is responsible for enforcing evaluation recommendations but two offices have related responsibilities. SISCONPES is a system designed by the National Planning Department to monitor responses to recommendation made in CONPES documents. The Directorate for Evaluation of Public Policies is responsible for promoting evaluation and supporting the design and implementation of evaluations recommended in CONPES documents. Those responsible for implementing and enforcing CONPES documents and other relevant legislation are primarily accountable to the National Planning Department's Director.

Colombia uses a mixed model approach to evaluation processes. Instead of the government being responsible for the implementation of evaluations, independent evaluators are responsible for implementing the evaluations designed within the government. Approximately 98% of evaluations are commissioned by SINERGIA and 2% are undertaken in-house. Importantly, the design of the evaluations (including the objectives, research questions and methods guidance) is mostly performed in-house by SINERGIA. The selection process for external evaluations is by competitive commissioning: almost every consultancy firm with some experience in evaluation is eligible to undertake evaluations. Individual consultants may be eligible to perform small evaluations or rapid appraisals.

Decision makers are involved throughout the processes of selecting, designing, commissioning, implementing and reporting. All evaluations led by SINERGIA also include an evaluation committee and the participation of decision makers from the programme or policy which is being evaluated is required. Typically the participants are people delegated by the relevant government ministers. Decision makers also participate via CONPES and the Intersectoral Evaluation Committee in the selection of programmes for evaluation, in the reporting of evaluations, and in analysing what actions to take in light of the evaluation results and recommendations.

SINERGIA also undertakes an intensive information dissemination process within the government and among programme stakeholders during each evaluation. This includes:

- A first-stage revision and discussion with technical staff from the units involved as well as the evaluation group from the Directorate for Evaluation of Public Policies

- A second presentation with managers and the heads of the units of the programme under evaluation, the Directorate for Evaluation of Public Policies, and the sector directorate of the National Planning Department. This presentation emphasises the evaluation findings and recommendations
- A third presentation stage with the heads of the agencies, the respective Minister, the Ministry of Finance budget director, the General Director of the National Planning Department, and staff

Following this, it is intended that an improvement plan is agreed between the programmes and Directorate for Evaluation of Public Policies and that compliance of commitments is followed-up by Directorate for Evaluation of Public Policies. The latter is currently being implemented and a technological system for monitoring of compliance with commitments derived from recommendations is currently under development. Although the results of evaluations are important for the National Planning Department, the Ministries and related institutions are responsible for making decisions based on the findings of evaluations.

Existing evaluation legislation is not linked directly to budgetary, financing or auditing processes. But guaranteeing that evaluations are of a high standard helps to ensure that the results from evaluations are used by decision makers. Legitimacy among decision makers, evaluation design relevance and feasibility, evaluation quality, and stakeholder participation are crucial determinants of evaluation effectiveness.

Legislation in Colombia does not require the active dissemination of evaluation results or communication with decision makers, but dissemination is required within SINERGIA's standard procedures. Making the results of evaluations publicly available is also not legally required. However, as most evaluations are commissioned and implemented using public funding, information and results can be accessed, as required, using Colombia's information access laws.

Attempts to increase accountability in Colombia have also included: the creation of partnerships with civil society organisations in order to disseminate monitoring and evaluation information; the production of programmes about government performance on television and radio stations; and contracting sectoral experts to analyse and report on government performance [2].

External funding continues to play an important role in evaluation funding but national government funding is increasing. Since 2002, at least two-thirds of the money spent on SINERGIA (donated by the Inter-American Development Bank, the World Bank and the government itself) has been spent on evaluations [2]. Currently, almost US$8 million in funding is made available annually. In 2010, the cost of evaluations was US$7 million: approximately US$5 million of this total was provided by international sources.

Effects of the legislation

The effects of the CONPES recommendations and other legislation used to establish the SINERGIA system have not yet been evaluated. However, SINERGIA – the national system for monitoring and evaluation – has been, and the evaluation included an assessment of SINERGIA's processes, the perceptions of stakeholders, and the effects of the SINERGIA's evaluation system. No legal challenges have been made thus far either to specific evaluations or to the system as a whole.

Approximately 30 evaluations are led by SINERGIA each year, and estimates suggest that 6% of the government spending on programmes that could potentially be evaluated is, in fact, evaluated. Currently, efforts are being made to estimate the proportion of evaluation recommendations that have been implemented.

Advocates and critics

While Colombia's monitoring and evaluation system receives general public support and approval, the independence and credibility of SINERGIA has been questioned by certain groups in academia and the private sector. Guaranteeing the objectivity of evaluations, they contend, is not possible given that SINERGIA remains under the control of the executive [2,3]. Nevertheless, despite this criticism, evaluations continue to be implemented by external firms in the private sector.

Strengths and weaknesses of the legislation

The government of Colombia has achieved considerable success in the creation and strengthening of SINERGIA. It remains one of the strongest governmental monitoring and evaluation systems in Latin America in terms of the range of monitoring and evaluation it undertakes, its overall credibility, and its utilisation and much of this progress was achieved since 2002, following the election of a reformist president. The creation of the SISMEG system to monitor the progress being made towards the achievement of the nation's Presidential Goals has been notable. And the intensive use of this system by the Presidency and the National Planning Department and the ambitious agenda of the evaluation system are also impressive.

A CONPES policy document on SINERGIA (#3294) lists four challenges facing the current monitoring and evaluation system:

- The lack of a single, clear conceptual framework
- The need to clarify the roles and responsibilities of the organisations supporting SINERGIA
- The absence of clear links between evaluation and planning processes and budgeting processes, and
- Problems with the availability and frequency of data and with data quality control

References

1. Briceño B, Gaarder MM. Institutionalizing evaluation: Review of International Experience. Egypt: International Initiative for Impact Evaluation, 2009.
2. Independent evaluation group: A Diagnosis of Colombia's National M&E System, SINERGIA. ECD working paper series no. 17. World Bank. 2007
3. Castro MF. Insider Insights: Building a Results-Based Management and Evaluation System in Colombia. World Bank. 2008.

Case Report: The Mexican Laws for Social Development and Financial Responsibility, and General Guidelines for the Evaluation of Federal Programmes

Enactment of the legislation

In 2004, the Mexican Congress unanimously approved the General Law of Social Development. This legislation was intended to ensure the full exercise of social rights, the establishment of mechanisms for evaluating and monitoring social development policy, and the periodic measurement of poverty. To achieve these goals, Mexico established The National Council for the Evaluation of Social Development Policies (CONEVAL), a decentralised public organisation under the Secretariat of Social Development, and provided it with both technical and management autonomy.

The creation of CONEVAL, however, was also strongly influenced by political pressure from the opposition. In 2001, The Technical Committee for the Measurement of Poverty had been formed as the result of an initiative by the Secretariat of Social Development (SEDESOL) to provide measurements of poverty that could be used as benchmarks for programmes and actions. Although SEDESOL successfully designed methods for measuring income poverty which were used at the national level, the organisation was also criticised for its perceived lack of objectivity and technical rigour. The creation of CONECAL was therefore important both to resolving existing problems and achieving new goals.

According to the 2001 Budget Law, all federal programmes subject to operational rules in Mexico are subject to annual external evaluation.

The General Law for Social Development (2004), the Federal Budget and Financial Responsibility Law (2006), and the Federation Expenditure Budget of 2007 conferred responsibilities for evaluation on three organisations, each with a specific scope:

- The National Council for the Evaluation of Social Development Policies (CONEVAL) has the power – based on the General Law for Social Development – to regulate and coordinate the evaluation of social development policies and programmes and to assess the compliance of programmes with their social objectives periodically.
- Together, the Secretariat of Finance and Public Credit (SHCP) and the Secretariat of Public Service (SFP) provide a performance evaluation system to identify the efficiency, economy, effectiveness and social impact of public expenditure, and in the case of the SFP, an evaluation of the performance and results of institutions. This system is based on the Federal Budget and Financial Responsibility Law.

The Federation Expenditure Budget passed by Congress during the 2007 fiscal year stipulated that CONEVAL, in association with the SHCP and the SFP, should issue

guidelines for evaluation, including a matrix of indicators, a monitoring and evaluation system, and strategic objectives for secretariats and agencies by the last working day of March that year. Accordingly, CONEVAL, SHCP and SFP issued the "General Guidelines for the Evaluation of Federal Programmes of the Federal Public Administration" in the Federation Official Journal (*Diario Oficial de la Federación*). The purpose of these Guidelines is to systematise the monitoring and evaluation of federal programmes in an integral manner such that the results obtained are able to inform decisions about public expenditures and measures and thereby contribute to making federal programmes more effective and efficient.

The enactment of the Guidelines was made possible by two key factors:

- The favourable political climate following the change of government in 2002 and the end of 70 years of single-party control. This enabled the modification of the structure of the Mexican Congress and the promotion of reforms directed at improving the effectiveness of government policies and programmes. Greater emphasis came to be placed on transparency, accountability, and results-based management.
- In 2006, rigorous evaluations of programmes such as *Oportunidades*, Social Provision of Milk (*Abasto Social de Leche*) and Firm Floor (*Piso Firme*) identified which programmes were working and which weren't. Programmes modifications were made to improve the role of evaluations evident in decision-making.

When designing the Guidelines, CONEVAL examined similar regulation schemes in other countries including:

- Chile: The implementation of the Administration Evaluation and Control System (*Sistema de Evaluación y Control de Gestión*) in the mid-90s consisted of a performance indicator system, an evaluation system and Administration Enhancement Programmes (*Programas de Mejoramiento de la Gestión*)
- Colombia: The creation of the National System for the Evaluation of Public Administration Results (*Sistema Nacional de Evaluación de Resultados de la Gestión Pública*), SINERGIA, was a response to a mandate required by the 1991 Constitution for the development of adequate monitoring and evaluation instruments in the public sector
- The USA: The 1993 Government Performance and Results Act (GPRA) was intended to lead to the development of an accountability framework for government agencies and the achievement of strategic results
- Australia: A reform process began in 1987, when programme evaluation was included as a component of a broad administration and budgetary reform for the first time. The intention was to strengthen departmental accountability in the application of public policies

Previous legislation in Mexico was also analysed during the development of the Guidelines. Prior to 2006, programme evaluations were regulated by the 2002 "Agreement by which the Minimum Requirements that shall be considered by Secretariats and Agencies of the Federal Public Administration in selecting academic and

research institutions or specialised bodies, national or international in nature, interested in performing the evaluation of government programme results subject to operation standards".

During the development process, the Inter-American Development Bank and other multilateral agencies provided technical assistance. More recently, the World Bank has provided information on experiences from other countries, including an analysis of aspects such as intergovernmental coordination, evaluation and monitoring instruments, and the link between evaluation and budgeting. In 2008, a year after the Guidelines were implemented, the World Bank issued a document titled 'Analysis Of The General Guidelines For The Evaluation Of Federal Programmes Of The Federal Public Administration' (*Análisis de los Lineamientos generales para la evaluación de los Programas Federales de la Administración Pública Federal*), from which CONEVAL adopted a series of recommendations to further improve implementation.

Since the initiation of the Guidelines, the Economic Commission for Latin America and the Caribbean (CEPAL) has provided technical support for the framework's logistical implementation. This has included methodological training for public servants and guidance on interpreting the matrices of indicators used by the programme operators.

Several additional documents were used during the design of the Guidelines (see Box).

Box. Documents used in the design of the Guidelines

Internal documents

- Proposal For The Design Of The Mexican Social Policy Institutional Evaluation and Monitoring System (Propuesta para el Diseño del Sistema Institucional de Evaluación y Monitoreo de Políticas Sociales en México). Issued as an internal CONEVAL document.
- Action Plan: Model To Strengthen The Federal Public Administration Results-Based Budgetary System (Modelo de fortalecimiento del sistema de presupuesto basado en resultados de la Administración Pública Federal). Issued as an internal SHCP document.

Documents about results-based management

- Administration And Results-Based Budget Control System. The Chilean Experience. (Sistema de Control de Gestión y Presupuesto por Resultados. La Experiencia Chilena.) Administration Control Division (División de Control de Gestión). Ministry of Treasure (Ministerio de Hacienda). Issued by the Government of Chile in 2003
- Handbook on Monitoring and Evaluating for Results. United Nations Development Programme. Issued by the Evaluation Office, 2002.

Documents used subsequent to the implementation of the Guidelines

- Analysis Of The General Guidelines For The Evaluation Of Federal Programmes Of The Federal Public Administration (Análisis de los Lineamientos generales para la evaluación de los Programas Federales de la Administración Pública Federal). A series of recommendations were used and adapted by CONEVAL to improve the implementation of the Guidelines. http://medusa.coneval.gob.mx/cmsconeval/rw/resource/coneval/quienes/2937.pdf
- Institute Bulletin No. 15: Logical Framework Methodology (Metodología del Marco Lógico). Latin American and Caribbean Institute for Economic and Social Planning (Instituto Latinoamericano y del Caribe de Planificación Económica y Social). Economic Commission for Latin America and the Caribbean (Comisión Económica para América Latina y el Caribe, CEPAL). http://medusa.coneval.gob.mx/cmsconeval/rw/resource/coneval/eval_mon/1322.pdf?view=true

Mexico attempted to establish a results-based management approach to evaluation through the implementation of the 1995-2000 Public Administration Modernization Programme (*Programa de Modernización de la Administración Pública 1995-2000*). This included the measurement and evaluation of public administration within sub-programmes. However, very little documentation or evidence survives from these initiatives and they therefore had limited impact upon the formation of the Guidelines. In 2002, the 'Agreement By Which The Minimum Requirements That Shall Be Considered By Secretariats And Agencies Of The Federal Public Administration In Selecting Academic And Research Institutions Or Specialized Bodies' (*Acuerdo por el que se expiden los Requisitos Mínimos que las dependencias y entidades de la Administración Pública…*) established the first annual evaluation scheme in Mexico. However, evidence related to the implementation process was similarly scarce and the extent to which evaluations were used between 2002 and 2006 is difficult to determine.

Legislation or regulations that facilitated the development and implementation of the 2007 Guidelines included:

- The General Law for Social Development (2004) and the Federal Budget and Financial Responsibility Law (2006) helped in the creation of standardised regulations
- The 2002 SHCP and SFP issued an agreement which established compulsory minimum requirements for evaluations: the hiring of external evaluators and the diffusion of information related to the evaluations. These proposals were used – and strengthened further – in the Guidelines
- The Federal Transparency and Access to Government Public Information Law (*Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental*) published in 2002, was created by the Federal Institute for Access to Information (*Instituto Federal de Acceso a la Información*, IFAI). This law guarantees the right of access to government public information. These principals and recommendations were used in the Guidelines to ensure greater transparency

Other legislation has restricted the regulatory-administrative commissioning process for evaluations. Secretariats and agencies, for example, must observe the Law of Procurement, Leases and Public Sector Services (*Ley de Adquisiciones, Arrendamientos y Servicios del Sector Público*). This law has occasionally prevented contracts being awarded to the best technical proposal because of economic considerations. The awarding of contracts can also be delayed.

Lessons that can be learnt by others considering the enactment of similar legislation or regulations include the importance of:

- Linking evaluation and monitoring objectives to other complementary government initiatives and institutions to create synergies in budgetary processes, accountability and transparency. This helps to create an integrated rather than overly-regulated system

- Seeking consensus and political support to promote reforms that foster the generation and use of indicators and evaluations, both for budgetary processes and for the improvement of public programmes
- Ensuring that reforms to implement a monitoring and evaluation system go hand in hand with administrative reforms. This enables those responsible for monitoring and evaluation to respond to the information needs of those making decisions and to link monitoring and evaluation to decision-making
- Building and continually improving monitoring and evaluation systems with reliable and objective information will give the system credibility

Scope of the legislation

The Guidelines require all federal secretariats and agencies to provide a diagnosis justifying the creation of new federal programmes based on their proposed contribution of these programmes to the strategic objectives of the secretariat or agencies, and to generate a matrix of indicators. All new programmes must undertake a 'design evaluation' within the first year (including revision of a logic framework and basic elements of design, such as the definition of the problem, the population, the relationship of the objectives and products, and possible duplication with other programmes). The relevant information is then delivered to either the SHCP or CONEVAL.

The Guidelines require impact evaluations for all programmes whether they are new or not. However, impact evaluations, given their technical rigour, information requirements, and cost, cannot always be implemented for all programmes. To this end, The Annual Evaluation Programme (*Programa Anual de Evaluación* or 'PAE') determines which programmes require impact evaluations, based on a consideration of the budget, size and relevance. A viability analysis is also undertaken by those wishing to conduct an impact evaluation and these may not be conducted if the results of the viability analysis are unfavourable.

The PAE, introduced in 2007, is defined jointly by CONEVAL, SHCP, and SFP and is a planning tool formalised in the General Guidelines for the Evaluation of Federal Programs. The PAE can mandate evaluations of consistency as well as the evaluation of results, design, indicators, processes, impacts and of specific strategies [2].

CONEVAL's scope covers all federal programmes in sectors related to social development, including health, education, the economy, agriculture and husbandry, the environment, agricultural reform, labour and social provision, social development, and science and technology. The application of the Guidelines to programmes unrelated to social development is the responsibility of the SHCP.

The Guidelines do not apply to state government programmes. However, CONEVAL is mandated to provide support at a state and municipal level in matters related to poverty evaluation. It therefore intends to replicate Guideline-based evaluations at the subnational level.

Together with other planning and budgetary reforms, Mexico's monitoring and evaluation system aims to establish *results*-based financing and management rather than placing an emphasis on *procedures* when deciding how to allocate public resources. This involves:

- Aligning national planning with the strategic objectives of federal secretariats and agencies and the goals of federal programmes
- Establishing results and management indicators
- Taking results from different types of evaluation into account (including design, indicators, processes, impacts and specific evaluations) in decisions about the allocation of public resources
- Simple budgetary standards which make it easier for secretariats and agencies to use resources efficiently

Characteristics of the legislation

All federal secretariats and agencies are required to adhere to the Guidelines and must use the monitoring and evaluation instruments prescribed. The progress of each programme is measured via the results and management indicators and these are then fed into the SHCP's internal evaluation system. Legislation requires that evaluation information related to government programmes be made public, but these programmes are not required to act on recommendations. Secretariats and agencies are legally required to publish data from commissioned evaluations on their websites in order to foster transparency and to ensure that evaluation processes are pursuant with the law. This data includes: information about external evaluators, information about the contract, the type of evaluation commissioned, the field information collected, the databases used, data collection instruments, and programme costs. CONEVAL also releases the results of its commissioned evaluations via its own publications.

The Guidelines establish the ex-ante development of impact evaluations. A technical review approved by CONEVAL, SHCP or SFP helps to ensure that an evaluation's objectives are well-defined, while the technical follow-ups undertaken by CONEVAL, SHCP or the SFP help to ensure their quality.

The Guidelines have established minimum criteria for some types of evaluation although they do not specify evaluation methods. For impact evaluations, methods are determined through feasibility analysis and through the terms of reference. In addition, some terms of reference issued by CONEVAL must be adhered to by federal secretariats and agencies. Other forms of evaluation may require more elaborate terms of reference. All such terms may be commented on by CONEVAL, SHCP and SFP. At present, CONEVAL provides terms of reference and guidelines for approximately 130 of the federal programmes under its mandate.

CONEVAL, SHCP and SFP coordinate or undertake the evaluation of programmes and policies, each within the scope of their particular authority and, thus far, all such

evaluations have been undertaken by external evaluators. The Annual Evaluation Programme establishes which of these organisations will coordinate or commissioned these programmes. The selection process must ensure transparency, equal opportunity and competitive conditions, pursuant to the Law of Procurement, Leases and Public Sector Services.

The Annual Evaluation Programme determines which evaluations will be funded by the Council and which will be funded by the Secretariats and Agencies. Of the 106 evaluations of consistency and results that the Council coordinated in 2007, it was possible to determine the cost of 102. The total cost was 71,034,321 Mexican pesos (US$5.9 million), which is less than 1% of the total budget allocated to the federal programmes evaluated that year; eight of these evaluations were funded by the Council. From 2008, the Council has provided resources for external evaluations, such as performance-specific evaluations. If other forms of evaluation are required by the secretariats and agencies, these are paid for from their own resources.

The Guidelines also identify factors that must be taken into account when selecting external evaluators, including: proof of legal incorporation, proof of experience in evaluation at the national or international level, and an executive brief for the proposed evaluation. According to the General Law for Social Development, academic and scientific research institutions and non-profit organisations may act as external evaluators. Further, external evaluators must be academic or social organisations that are independent of the government and have prior experience related to the relevant subject.

Direct links with the budgetary allocation are not defined by the Guidelines. However, evaluations provide input and are used as reference points that are promoted in the budgetary process. Results and management indicators and performance-specific evaluations feed into the Performance Evaluation System (for which SHCP is responsible), the aim of which is to provide information for budgetary decision-making.

CONEVAL is required to deliver the results of social programme evaluations to the Social Development Commissions of Congress, as well as to the SHCP, SFP and the Federal Auditing Office (*Auditoría Superior de la Federación*). All evaluations conducted externally must be delivered to the Chamber of Representatives' Budgetary and Public Account Commission (*Comisión de Presupuesto y Cuenta Pública*), the SHCP, SFP and the Federal Auditing Office (*Auditoría Superior de la Federación*). Information from external evaluations of social programmes is sent by CONEVAL to SHCP in a document entitled 'Considerations for the Budgetary Process' ('*Consideraciones para el Proceso Presupuestario*').

Following this, CONEVAL prepares a document based on the results of external evaluations, entitled 'Follow-up Report on the Aspects of Federal Programmes that Can Be Improved' ('*Informe de seguimiento a los Aspectos Susceptibles de Mejora de Programas Federales*'). Those responsible for each programme then prepare a

work programme derived from the Follow-up Reports, including commitments to recommendations that they accept. CONEVAL also prepares the 'Evaluation Report on Social Development Policy' ('*Informe de Evaluación de la Política de Desarrollo Social*'), which establishes recommendations addressed to different decision-makers such as the Chamber of Representatives, the Inter-Ministry Social Development Commission, the National Social Development Commission, and social programmes. CONEVAL also uses forums within the legislative and the executive branches of government and within civil society to disseminate evaluation results.

Effects of the legislation

In 2008, the World Bank undertook a study of the implementation of Mexico's Guidelines. This study analysed the first phase of Guideline implementation and was based on an analysis of the use of indicator matrices and of the results of the consistency evaluations. The Guidelines were found to have positively contributed to evaluation in Mexico by:

- Creating a common results-based language and system within the Federal Public Administration
- Establishing the beginning of a learning process among public servants about monitoring and evaluation
- Providing a standardised menu of monitoring and evaluation instruments to achieve better results, both for the secretariats and for the programmes

Mexico implemented a number of changes based on the recommendations of the report, which had recommended the following actions:

- Enhance the timely diffusion of the results of public programmes and policy evaluations
  - Publish the external evaluation reports on CONEVAL's website
  - Advertise the availability of evaluations on CONEVAL's website and in national newspapers
- Establish an inter-institutional committee to promote scheduled interaction and communication between the parties involved with the Guidelines.
  - CONEVAL and SHCP have jointly developed criteria, manuals and technical notes for the formal approval of training and technical assistance related to the matrices of indicators
  - CONEVAL and SHCP have issued general criteria for operational rules with the matrix of indicators
  - CONEVAL, SHCP and SFP have published the 'Mechanisms for the Follow-up of Aspects Susceptible to Improvement in External Evaluations' ('*Mecanismos de Seguimiento a los Aspectos Susceptibles de Mejora de las Evaluaciones Externas*')
- Create an evaluation register
  - CONEVAL now registers external evaluators or potential evaluators who wish to take part in the Evaluating Bodies Directory on its website

From 2007 to 2010, 477 external evaluations of various kinds were undertaken. Performance-Specific Evaluations (the instrument used to evaluate federal programmes annually) were used in 127 of the 158 federal programmes undertaken during the 2009-2010 term – i.e. approximately 80% of those that could potentially be evaluated.

Examples of success and failure of the legislation

No mechanisms are available to identify all the actions undertaken based on evaluations, but these include the following examples:

- The transformation of the ProTree Programme (*Programa ProÁrbol*) to incorporate new intervention schemes in order to increase its effectiveness. These changes were based on the recommendations to the 2007 consistency and results evaluations
- The cancellation of the Social Security First Job Programme (*Programa Primer Empleo del IMSS*). The design evaluation found that the programme did not comply with the purposes for which had been created
- Recommendations made by the 2008 Evaluation Report on Social Development Policy led to the creation of the Five Steps For Your Health to Live Better Programme (*Cinco Pasos para su Salud para Vivir Mejor*), the expansion of the National Infrastructure Programme (*Programa Nacional de Infraestructura*), the Social Security period of coverage, the extension of the Temporary Job Programme (*Programa de Empleo Temporal*) to urban areas, and the restructuring of the Procampo operating rules (these establish the way in which a programme will operate, describe its objectives, population, specific support, the actions that will be undertaken, and the conditions under which people will receive the benefit)
- Using information from consistency and results evaluations, performance-specific evaluations, and the analysis of the progressiveness of the programmes (i.e. the relationship between expenditures and income levels) to inform the 2010 budgetary allocation
- The Federation Expenditure Budget for fiscal year 2010 stated that social programmes (Branch 20) must be focused on populations in which poverty, vulnerability, marginalisation and social gaps are defined in accordance with the official standards recognised by SEDESOL, the National Council of Population (CONAPO) and CONEVAL assessments.
- The updating of SEDESOL's diagnostics of the following social programmes: Habitat (*Hábitat*), Daycare Centres to Support Working Mothers, (*Estancias Infantiles para Apoyar a Madres Trabajadoras*), Production Options (*Opciones Productivas*), Attention for Agricultural Laborers, (*Atención a Jornaleros Agrícolas*), Rescue of Public Spaces (Rescate *de Espacios Públicos*), Rural Supply (*Abasto Rural*), Temporary Employment (*Empleo Temporal*), National Fund for Handicraft Development (*Fondo Nacional para el Fomento de las Artesanías*), Support for Instances of Woman in Federal States (*Apoyo a las Instancias de Mujeres en las Entidades Federativas*), Rural Housing (*Vivienda Rural*) and Savings and Housing Subsidy Your House (*Ahorro y Subsidio para la Vivienda Tu Casa*) .

- The modification of the operational rules of the following social programmes based on the findings from the Performance-Specific Evaluations: Rural Housing (*Vivienda Rural*), Support Fund for Micro, Small and Medium Business (*Fondo de Apoyo para la Micro, Pequeña y Mediana Empresa*), Attention to Adult Education Demand (*Atención a la Demanda de Educación para Adultos*), Clean Water (*Agua Limpia*) and Environmental Institutional Development Program (*Programa de Desarrollo Institucional Ambiental*),
- 491 programmes thus far have used results and management indicators developed as a result of the ongoing coordination between CONEVAL and the secretariats and agencies. 158 of these were social development programmes.

Programme managers continue to produce evaluations and are committed to complying with the 'Mechanism for The Follow-Up of Aspects Susceptible to Improvement'. The way in which evaluations are conceptualised has changed significantly: in the past they were seen as an obligation and more as an inquiry than an enquiry. Commitments to linking evaluations to improvements in management were not always strong. Evaluation is now seen as a useful tool for improving programme operation: more than 70% of the programmes evaluated follow-up on their evaluation results. CONEVAL's mechanism for conducting follow-ups to evaluation results consists of an analysis of the viability of the programmes recommendations and a programme of works. These establish the actions that programmes need to take in order to address the recommendations made, including identifying who is responsible and the time frame required.

The changes proposed to social development policies that involve more than one actor and that require institutional coordination between secretariats and agencies or between different government levels have generally been limited. In such cases, the diversity of actors may be an obstacle to improvement and the evaluation objectives in such instances may not fully align with, for example, health sector priorities.

Advancements have been made in the enactment of social development laws at a subnational level, including requirements for evaluation. In some cases, institutions similar to CONEVAL have been established. More than ten states in Mexico now have legislation related to social development, and their approaches are very similar to those devised originally by CONEVAL. Some states have organisation structures for evaluation which are also similar to CONEVAL; in others, evaluation is undertaken within specific units focusing on social development.

Advocates and critics

Representatives from CONEVAL, SFP and SHCP, programme managers who want to show that their programmes have provided positive results (for technical or political reasons) and legislators with an interest in social development issues, are among the strongest advocates of the current evaluation system.

Different actors in the executive and legislative branches of government have proposed modifications, both to the regulations administered by CONEVAL, and to the authority itself. Some critics of the Social Development Commission, for example, have argued that CONEVAL lacks the instruments and authority required to make the Guideline-based recommendations mandatory. But, thus far, no court challenges have been launched again the Guidelines and, overall, the responses to the Guidelines have been positive.

Important evaluation issues, including the results of the social programmes, have been discussed and reviewed in the media. The fact that the legislation requires the publication of evaluation results has also contributed to raising the level of public debate by providing data measuring public sector performance.

Strengths and weaknesses of the legislation

The main strengths of the legislation include:

- Providing homogenous evaluation and monitoring criteria for all secretariats and agencies in the Federal Public Administration, thus contributing to an integrated information scheme
- Establishing links between the monitoring and evaluation system and the central public decision-making processes (such as budgetary, strategic planning and institutional management), irrespective of which parties are responsible for executing those tasks
- Actively involving people in monitoring and evaluation actions

CONEVAL has made several changes to the Guideline to address the implementation weaknesses identified by the World Bank [3]. However, some weaknesses that were identified required decisions beyond CONEVAL's authority. These included:

- Requirements under the Law for Acquisitions, Leasing and Services by the Public Sector to limit how contracts for evaluations are awarded
- The need to make recommendations derived from evaluations binding. In other words, to ensure that there are mechanisms to ensure that the results of evaluations are linked to decisions about the allocation of resources and the modification or suspension of programmes and policies as needed.
- The need to improve administrative registers of the secretariats and agencies so that they are used more effectively as a source of information for evaluations, and thereby raise the quality of evaluations.

An additional weaknesses in the evaluation process is the need for better legal congruence and regulation for external contracts. This will ensure better clarity in the evaluation contracting process. Currently, evaluation contracting takes place pursuant to the Law for Acquisitions, Leasing and Services by the Public Sector.

References

1. Briceño B, Gaarder MM. Institutionalizing evaluation: Review of International Experience. Egypt: International Initiative for Impact Evaluation, 2009.
2. Gaarder MM, Briceño B. Institutionalization of Government Evaluation: Balancing Trade Offs. International Initiative for Impact Evaluation, 2010.
3. Unidad de Pobreza y Género Reducción de la Pobreza y Gestión Económica Región de América Latina y el Caribe. México: Análisis de los Lineamientos generales para la evaluación de los Programas Federales de la Administración Pública Federal. http://medusa.coneval.gob.mx/cmsconeval/rw/resource/coneval/quienes/2937.pdf

Case Report: Evaluation of Educational Programmes in the USA

Enactment

The No Child Left Behind Act [1] was enacted in 2001, the Education Sciences Reform Act [2] in 2002, and the Individuals with Disabilities Education Improvement Act [3] in 2004. The first two of these Acts were amendments to previous legislation ('No Child' was an amendment to the Elementary and Secondary Education Act of 1965). It can therefore be argued that the implementation of these laws was not driven by special circumstances, but by a desire to improve previous legislation. Both The No Child Left Behind Act and the Individuals with Disabilities Education Improvement Act were initiated by the Bush administration, with the involvement and support of Members of Congress and Senators. The Bush administration was less involved in the Education Sciences Reform Act, which was promoted by a Member of Congress from Delaware.

Professional associations lobbied for and against particular aspects of the 'No Child' and 'Disabilities Education Improvement' Acts, including the American Educational Research Association, the American Psychological Association, and the Knowledge Alliance. The American Educational Research Association was opposed to provisions of the legislation which strengthened the definition of rigorous research to refer to randomised trials. At the same time, they were in favour of provisions of the legislation that strengthened the independence of the Institute of Education Sciences (e.g. the independence of the publication authority and a six year term limit for the director). International organisations did not play any role in the development of these Acts.

These laws and amendments were made possible with the support of Members of Congress as well as the Bush administration's strong focus on education; no significant legislative barriers were apparent. Education research had been in the spotlight in the USA owing both to long-standing concerns and disappointment related to the quality of education, and because of the importance of good education to the economy and the future of America.

Opposition to the legislation itself was minimal because it remained unclear how involved and committed to using rigorous methods the Institute of Education Sciences would be. Subsequently, there has been substantial opposition to the use of rigorous evaluation.

Scope of the legislation

Research evidence is used in American federal education programmes, sometimes with the launch of a programme and sometimes at the end of a programme. However, the use of systematic reviews is not required. Evaluations are conducted during

the period of programme authorisation (typically every 4-5 years) to inform reauthorisation by Congress.

Big budget programmes nearly always have mandated evaluations or budgets set aside for national activities to support evaluations. Smaller programmes typically do not have legislatively mandated evaluations and do not generate sufficient administrative funds to make optional evaluations affordable. The legislation relates to the evaluation of different programmes within the education sector and different levels of government are involved: sometimes recipients of funding are required to participate in federally-conducted evaluations, whereas in other cases, the recipients are required to conduct the evaluations themselves.

The Institute of Education Sciences has overarching responsibility according to the Education Sciences Reform Act for evaluating education programmes for the disadvantaged (as defined according to the Elementary and Secondary Education Act). It is also often designated in legislation passed by Congress as the agency responsible for the evaluation and authorisation of particular education programmes. For many other education programmes, a case-by-case assessment is made for each programme to determine the specific manner in which it is to be evaluated and who will be responsible. An informal process in the Department is used to propose an annual evaluation plan, and related negotiations continue until the plan is finalised. The Institute of Education Sciences has been responsible in recent years for all *evaluations* of programme impact, whereas the Office of Planning, Evaluation, and Policy Development has frequently been responsible for evaluations of programme *implementation*.

Characteristics of the legislation

Legislation requiring programme evaluation is binding for the U.S. Department of Education and can be enforced by budget cuts. However, in practice this does not occur because the legislation is typically adhered to. In instances where adherence is not possible, the responsible committees in Congress are contacted and permission is sought to deviate from legislative requirements. Sometimes Congress specifies particular evaluation methods or particular questions that need to be answered before a programme evaluation is authorised; in other cases, the process is less prescriptive. The actual content details of an evaluation are descriptions of the work to be accomplished and are a joint product generated by the U.S. Department of Education and external contractors competing for the work, who are asked to prepare responses detailing how they will translate the Department's general requirements into action. These contractors (who are outside of government), collect and analyse data, and write draft reports. The office within the federal agency that commissioned the work will decide ultimately how the information is to be released. In some cases, the unit commissioning an evaluation will not be independent from the entity which is being evaluated during the reporting process. In such circumstances, it cannot in-

dependently release the results of an evaluation. However, the Institute of Education Sciences has been legally granted such independence.

Legislation does not stipulate how evaluation results are to be used by decision makers. The Institute of Education Sciences is mandated to make evaluation results available publicly and to make data available to external researchers wishing to undertake their own evaluation (section 184 and 186 of the Education Sciences Reform Act [2]). There is no independent funding for evaluation available in the Department of Education and all such funding (US$70 million annually) comes from funds set aside specifically for evaluation in the authorisation of other programmes. The budget allocated for evaluation is small compared to the Department of Education's total US$60 billion budget.

Effects of the legislation

The legislation itself has been evaluated and recommendations made. One recommendation is that all evaluations should be the responsibility of the Institute of Education Sciences rather than sometimes being the responsibility of offices which are not independent of the political leadership of the Department of Education. Prior to the reauthorisation of the Education Sciences Reform Act, the Obama administration took the constructive step of clearly demarcating the responsibilities for evaluation between the Institute of Education Sciences and others responsible for more short-term assessments.

Evaluations by the Institute of Education Sciences do not carry policy recommendations with them. The purpose of its evaluations is to describe only what is found: how the results are responded to is left up to policymakers and to the broader political process. According to The Institute of Education Sciences, the inclusion of policy recommendations in evaluation reports introduces values and political standpoints that may undermine the perceived objectivity of an evaluation. Although IES reports may contain recommendations for further research and may summarise the results of the related research, they do not contain recommendations for legislative or administrative actions.

The evaluation of small programmes is limited by a lack of funding. Administrations frequently try to rid themselves of these as they are often viewed as bothersome. This lack of evidence means that decisions on the continuation, termination, or revision of small programmes remain largely uninformed by information about their effectiveness.

All large programmes are subjected to some type of evaluation.

Advocates and critics

Advocates of evaluation are frequently those who have pushed the relevant laws forward. Critics of evaluation findings are often those who have been working in a programme which has been found to be ineffective. Typical responses to negative findings include disbelief and a certainty that key issues have not in fact been measured. Criticism is often primarily focused on the findings of the evaluations but may also extend to being post-hoc criticism of the entire evaluation process.

The organised research and evaluation community in the USA (e.g. the American Evaluation Association) may also be critical. Many of its members object to the use of randomised trials and could probably not be described as keen advocates of impact evaluation. At times, they have advocated against stronger legislation for evaluations. Many believe that such methods yield similar results to other strategies and that the government should not tie the hands of evaluators by suggesting that one particular approach may be better than others.

Media coverage of evaluation results is typically confined to instances in which the results may be interesting to the public, but the evaluation processes are seldom reviewed.

Strengths and weaknesses

The Education Sciences Reform Act was a significant improvement upon previous legislation and a model for governmental involvement in the process of evaluation. It provides for the appointment of a director for a six-year fixed term. In the USA, executive branch employees at this level serve 'at the pleasure' of the president (mostly this means they are working for the cabinet level official to whom they report). A fixed-term appointment allows the director to make decisions without fear of immediate dismissal and sends a strong signal that the office for which the official is responsible must be non-partisan. The Education Sciences Reform Act also provides for an independent publication authority, which means that the evaluation reports are not subject to the review and approval of political officials who may prefer *not* to have a programme they're associated with to be found wanting by a federal evaluation. Independent peer review also provides for the control of quality and independence and the timely release of all data. All staff are hired at the discretion of the director – this too provides independence from external pressures to fill such positions with people who may otherwise support an administration's political and policy positions. A board of members from outside of government oversees the work of the Institute of Education Sciences and makes periodic reports to Congress and the public on the Institute's functioning. Many regard the Education Sciences Reform Act as a model for evaluation legislation.

Many researchers (mostly university-based) compete for research, and over 1,000 outstanding grants have been awarded. Support is also provided to doctorate-level training programmes and the legislation has created capacity that would not otherwise have been available. The quality of education research funded by the Institute

of Education Sciences in America now rivals that of the social and behavioural research funded by the National Science Foundation and the National Institutes of Health. Evaluations results have also had significant impacts on the nature of federal programmes.

A weakness is that there is no comparable local and state capacity and dissemination to local and state programmes has not been developed. There is a need to support the uptake of evaluation results by practitioners and policymakers at the local and state levels. There is also a lack of independent funding for evaluations.

The current laws also have specific limitations. The No Child Left Behind Act, for example, requires the use of evidence-based programmes even in instances in which these are unavailable.

Another problem is that Congress typically requires evaluation of major federal programmes or initiatives with lots of specific questions. The questions are often poorly formulated and difficult to answer.

Many recommendations have been made for the revision of the Education Sciences Reform Act, including changing the start of the director's term so that it does not coincide with a change of administration. This will help to keep such appointments outside the heavily-politicised process of appointing senior positions during a change of government.

References

1. The No Child Left Behind Act of 2001, (Public Law 107-110). http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf
2. Education Sciences Reform Act of 2002. http://ies.ed.gov/pdf/PL107-279.pdf
3. The Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004) http://www.copaa.org/pdf/IDEA2004.pdf

Appendix 6. Excluded examples

Country	Legislation/regulation/Policy	Exclusion criteria
Australia	Cabinet evaluation strategy (1988 to 1997)[9]	Insufficient information[10]
Australia	Environmental Impact Assessment	EIA[11]
Australia	Family Impact Statement	RIA[11]
Australia	Health Technology Assessment	HTA[11]
Australia	Regulatory Impact Statement (1995)	RIA[11]
Canada	Cabinet directive on streamlining regulation (2005)	RIA[11]
Canada	Regulatory Impact Analysis (1995)	RIA[11]
Canada	Research Evidence Tool (2010)	RIA[11]
England	HTA by the National Institute for Health and Clinical Excellence (NICE)	HTA[11]
European Union	Environmental Impact Assessment directive (1985, later amendments)	EIA[11]
Millennium Challenge Corporation	Policy for monitoring and evaluation of threshold programmes (2009)	Not mandatory for all programmes[12]
Norway	Veileder til gjennomføring av evalueringer (2005)	Not mandatory
Philippines	Republic Act nr 78/75 (National Health Insurance Act of 1995)	Evaluation of a specific programme[13]
South Africa	Framework for Government Wide Monitoring and Evaluation (2007)	Not mandatory
Spain	Evaluation at AEVAL (2006 Law of State Agencies)	Not mandatory
Switzerland	Swiss Federal Constitution, Article 170 (2000)[14 15]	Not mandatory evaluation[16]
Thailand	Article 67 of 2007 Constitution	HIA[11]
Thailand	Article 10 and 11 of the 2007 National Health Act	HIA, based upon request
United Kingdom	The Magenta Book: Guidance for evaluation (2011)	Not mandatory for all programmes[12]
United Nations	UNDP evaluation policy (2006)	Not mandatory
United Nations	UNIDO evaluation policy (2006)	Not mandatory for all programmes[12]
United Nations	WFP evaluation policy (2008)	Not mandatory for all programmes[12]
United Nations	UNIFEM evaluation policy (2009)	Not mandatory for all programmes[12]
United Nations	IFAD evaluation policy (2003)	Not mandatory
United Nations	IAEA evaluation policy (2002)	Not mandatory
United Nations	GEF evaluation policy (2006)	Not mandatory

[9] Mackay K. Two generations of performance evaluation and management system in Australia. World Bank, ECD Working Paper Series 11, 2004.

[10] In 1988 Cabinet agreed to an evaluation strategy that included formal requirements for departments to evaluate every programme every 3-5 years, prepare annual evaluation plans, and to include proposed arrangements for evaluation in new policy proposals. These formal requirements were eliminated in 1997 by a newly elected conservative Coalition government. Although there are published descriptions of this policy, due to the historical nature of this example, we were unable to obtain information about this policy from key informants, comparable to the examples included in this review.

[11] EIA = Environmental Impact Assessment; HIA = Health Impact Assessment; HTA = Health Technology Assessment; RIA = Regulatory Impact Assessment

[12] I.e. does not require routine (systematic and transparent) consideration of evaluation or the need for evaluation

[13] Requirements for evaluation of a single specific programme

[14] Widmer T, Neuenschwander P. Embedding evaluation in the Swiss federal administration: Purpose, institutional design and utilization. Evaluation 2004; 10:388-409.

[15] Jantz B, Veit S. Sunset legislation and better regulation: Empirical evidence from four countries. Gütersloh: Bertelsmann Stiftung, June 2010.

[16] Requires the Federal Parliament to ensure that the effectiveness of federal measures is examined, but does not specify the use of evaluation or details of institutionalisation. A central oversight body does not exist; instead, a multitude of institutions evaluates the impact of state activities, but there is no consensus about what evaluation is or should do.

Country	Legislation/regulation/Policy	Exclusion criteria
United Nations	World Bank Evaluation Policy (2005)	Not mandatory
USA	Recovery Act (2009)	Mandatory under financial preconditions
USA	Supplemental Appropriations Act (2009)	Not mandatory
USA	Second Chance Act (2008)	Not mandatory for all programmes[12]
USA	Higher education opportunity Act (2008)	Not mandatory for all programmes,[12] financial preconditions
USA	Office of Management and Budget initiatives (2009)	Not mandatory
USA	Patient Protection and Affordable Care Act of 2010	Not mandatory for all programmes[12]
USA	USAID Evaluation Policy (2011)[17]	Published after data collection was completed for this review

[17] USAID. Evaluation: Learning from Experience. USAID Evaluation Policy. Washington DC: USAID, 2011.

Acronyms

AEVAL	State Agency for the Evaluation of Public Policies and Quality of Services in Spain
CAD	Canadian Dollars
CONEVAL	National Council for the Evaluation of Social Development Policies in Mexico
CONPES	National Council for Economic and Social Policy in Colombia
DIPRES	National Budget Bureau in Chile
EIA	Environmental Impact Assessment
EMS	Expenditure Management System in Canada
GEF	Global Environment Facility
HIA	Health Impact Assessment
HTA	Health Technology Assessment
IAEA	International Atomic Energy Agency
IES	Institute of Education Sciences in the USA
IFAD	International Fund for Agricultural Development
INAHTA	International Network of Agencies for Health Technology Assessment
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
OECD	Organisation for Economic Cooperation and Development
PAE	Annual Evaluation Programme in Mexico
RIA	Regulatory Impact Assessment
SEDESOL	Secretariat of Social Development in Mexico
SFP	Secretariat of Public Service in Mexico
SHCP	Secretariat of Finance and Public Credit
SINERGIA	National System for Monitoring and Evaluation in Colombia
UNDP	United Nations Development Programme
UNIDO	United Nations Industrial Development Organization

UNIFEM	United Nations Development Fund for Women
USA	United States of America
USAID	United States Agency for International Development
WFP	World Food Programme
WHO	World Health Organization